A light-weight concept ontology for annotating digital music.

RAHMAN, Fazilatur.

Published version

RAHMAN, Fazilatur. (2010). A light-weight concept ontology for annotating digital music. Doctoral, Sheffield Hallam University (United Kingdom)..

A Light-weight Concept Ontology for Annotating Digital Music

Fazilatur Rahman

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

July 2010

Dedication

I remember the all-embracing kindness and the blessings of Almighty Allah who graciously has given me this extra-ordinary opportunity to complete this work by putting together all sorts of support and help from every corner I needed. I am ever grateful to HIM and pray to the omnipotent being to remove chances of errors and mistakes by me and make this effort successful.

I owe this endeavour to my late mother Manjaratun Nessa, whose prayers always stood by my side though she could not witness it. Blessed be her soul rests in peace.

Abstract

In the recent time, the digital music items on the internet have been evolving to an enormous information space where we try to find/locate the piece of information of our choice by means of search engine. The current trend of searching for music by means of music consumers' keywords/tags is unable to provide satisfactory search results; and search and retrieval of music may be potentially improved if music metadata is created from semantic information provided by association of end-users' tags with acoustic metadata which is easy to extract automatically from digital music items. Based on this observation, our research objective was to investigate how music producers may be able to annotate music against MPEG-7 description (with its acoustic metadata) to deliver meaningful search results. In addressing this question, we investigated the potential of multimedia ontologies to serve as backbone for annotating music items and prospective application scenarios of semantic technologies in the digital music industry. We achieved with our main contribution under this thesis is the first prototype of ***mpeg-7Music*** annotation ontology that establishes a mapping of end-users tags with MPEG-7 acoustic metadata as well as extends upper level multimedia ontologies with end-user tags. Additionally, we have developed a semi-automatic annotation tool to demonstrate the potential of the ***mpeg-7Music*** ontology to serve as light weight concept ontology for annotating digital music by music producers. The proposed ontology has been encoded in dominant semantic web ontology standard OWLI.O and provides a standard interoperable representation of the generated semantic metadata. Our innovations in designing the semantic annotation tool were focussed on supporting the music annotation vocabulary (i.e. the ***mpeg-7Music)*** in an attempt to turn the music metadata information space to a knowledgebase.

Acknowledgement

I received invaluable help and support to present my research in this thesis as a coherent piece of work. My heart-felt gratitude goes to my Director of Studies Dr. Jawed Siddiqi whose patient and regular supervision almost every week throughout last four years made me bring this work to its present shape. He has been a friend, philosopher and guide whom I could knock whenever I wanted to discuss anything I thought important. My special thanks go to Dr. Babak Akhgar for his informative advice, affectionate attention to support me by bringing any resource that could improve my understanding of research fields.

I am grateful to Dr. Nahum Korda and colleagues (Igor Czechak, Roman Dobjanski) at the Straight Technology for their priceless effort to provide me the chance to get hands-on experience on cutting-edge technologies.

My humble thanks to colleagues at Cultural Communications and Computing Reasearch Institute where in various capacity, I came in contact with Dr. Sally Atkinson, Dr. Christopher Bates, Dr. Simon Polovina, Lukasz Jopek, Dr. Andrew Dearden, Dr. Simeon Yates are to name a few who whole-heartedly extended their hands of support to bring my research to a reality. I would also like to thank staffs at Sheffield Hallam University's Research Office who either directly or indirectly came to my assistance and especially to Dr. Kathy Doherty. Fellow PhD students - Ja'far, Nasseer, Mohammed, Khalid do deserve my thanks and gratitude, they were ever ready to help whenever possible.

I wish to acknowledge selfless support from Dr. Mike Holcombe (of University of Sheffield) who encouraged me to commit to a direct PhD program like this one.

The names I have mentioned above represent a partial list of those to whom I am indebted to. There are so many of my teachers, mentors, friends, relatives who made an impact in shaping

my thinking, observations and of course my family, parents & siblings - where I have been raised to be the person I am today.

Finally, I feel proud to mention my family – my husband Shawkat and my son Krishn - for being patient about many evenings, weekends and holidays to allow me to research and write this thesis rather than spending time with them.

List of papers

1. Jawed Siddiqi, Nazaraf Shah, Babak Akhgar and Fazilatur Rahman. A System for Semantically Enhanced, Multifaceted, Collaborative Access: Requirements and Architecture. 4th International Conference on Information Technology: New Generations. ITNG 2007. April 2-4, 2007, Las Vegas, Nevada, USA
2. Jawed Siddiqi, Babak Akhgar, Fazilatur Rahman, Nazaraf Shah, Nahum Korda , Raphael Attias, Norbert Benamou, Andrade Maria Teresa , Judy Dori, Boas Hashavia. Collaborative Access through Semantically Enhanced Distributed & Diversified Cultural Content. The 2007 International Symposium on Collaborative Technologies and Systems (CTS 2007). May 21-25, 2007. Orlando, USA.
3. Jawed Siddiqi, Babak Akhgar, Fazilatur Rahman, Nazaraf Shah, Nahum Korda , Raphael Attias, Norbert Benamou, Andrade Maria Teresa , Judy Dori, Boas Hashavia. Semantic Modelling of Digital Multimedia. The 2007 International Conference on Semantic Web and Web Services (SWWS'07). Monte Carlo Resort, Las Vegas, Nevada, USA (June 25-28, 2007).
4. Fazilatur Rahman, Jawed I. Siddiqi, Nidhi Bindal, Laura E. Raileanu and Nastaran Fatemi. MOSAICA: A Knowledge Portal for Cultural Heritages. The 2008 International Conference on Information and Knowledge Engineering (IKE'08). Monte Carlo Resort, Las Vegas, Nevada, USA. July 14-17, 2008.

List of Figures

List of Tables

Table of Contents

1 - Introduction

1.1. Motivation and Background

We all want to find music the way we like it or want to utilize it. Music professionals (composers, musicians or song writers) are interested about musical score, arrangements and lyrics. Record companies dealing with sound recording, playback and marketing of musical artefacts are interested in categorizing music for commercial purposes. Ordinary listeners are keen to find, listen and share their favourite music and songs. The Worldwide Web/Internet has changed the music industry by making huge amount of music available to both music publishers and consumers including ordinary listeners or end users. As a result, we are faced with the desirable or otherwise challenge of how we can find the music of our choice.

As ordinary listeners, we want to know more about our favourite musical pieces so that we can easily and effectively find, retrieve and share it. Until now, we try to search music by filename, artist name, album title, musical style or genre - technically these are termed as syntactic metadata. These sales oriented metadata or categorization tags provide a very limited collection of vocabulary/keywords. Music search based on these syntactic metadata requires the search query to contain at least one keyword from that vocabulary and it must be an exact match. But, keywords used by ordinary music fans often contain diverse range of description of the song that differs from the publishers categorization tags. As a result, consumers are left with unsatisfactory search results.

From a technical perspective music is also expressed by another type of syntactic metadata which is actually audio signal level statistics and is only understood by machines but not by ordinary end users. As a result signal level metadata are not included in the end-users vocabulary. But, signal level metadata are often used to automatically detect musical style or acoustic properties to detect musical instruments, key or structure. Such an approach could be used to yield satisfactory search results but the acoustic metadata is not suitable to be used for search and tagging purpose as meaning of those metadata are not perceived by ordinary human users while searching/annotating songs of interest.

Moreover, a single song depicting a particular acoustic property or auditory perceptual features may be interpreted differently by different music users. For example, a sample description of a song may be formulated as:

"This song is **mellow** and **sad**. It features **airy vocals**. It is a song with a **slow tempo** and **rising melody** and with **low energy**."

The above description contains the bold-faced metadata (i.e. metadata from ordinary music users' vocabulary) that can be associated with signal level acoustic metadata. Depending on the users level of understanding of music, idea, mood and intention they may derive different meanings from the acoustic features. Same acoustic features may be interpreted differently by different users. This derivation of meaning is technically termed as semantic interpretation. Some users may interpret a piece of music as "My favourite song" or "Sounds like 70s" or "Romantic Music" where that piece may be derived from the same set of acoustic features using different relationship of acoustic properties and its interpretation.

Derivation of semantic interpretation from different relationship of acoustic properties has been tried by researchers e.g. in (Whitman 2004). But, simple mapping or interpretation of acoustic relationship is not enough to be used as a means for effective music search and retrieval as the efficiency of search and retrieval is dependent on the structured and meaningful mapping of the machine level statistics. The structured representation of machine level acoustic information of music (standardized by MPEG-7 (Martinez et al., 2002)) is not understood by most of the end users though those perceptual feature based representation could lead to finding of that music media easily. So, it would be very useful if we could enrich the semantic interpretation of musical pieces with those acoustic/perceptual features following standard representational techniques. At present, ordinary users metadata vocabulary to describe music is quite unstructured and is not defined with any standard interpretation mechanism that can be linked with acoustic metadata that can be automatically derived from music audio. This thesis presents a structured semantic metadata that creates a mapping between music consumers' metadata and the underlying music's acoustic metadata in order to enable music producers to annotate music in a meaningful way that is actually termed as 'semantic annotation of music'.

Now, the leading edge technologies proposed by Semantic Web Initiative pioneers the idea of creating meaningful & structured representation of the content of text, image, audio (music is a special type of audio) and video available in formats those will be under stood by both humans and machines. Musical contents categorized using structured semantic metadata will be easily retrievable by machines. However, semantic web technologies do not provide the detail of techniques on how to annotate music using the structured metadata vocabulary to generate categorized musical contents. The Web2.0 technologies (Damme et al., 2007) are somehow successful to create applications to engage/enable ordinary users to categorize contents but the generated annotations are not meant for machine processing. Manipulation of such annotations boils down to the problem of in-efficient search and retrieval. Given this background of available technological options it may be able to bridge these two technologies to build a solution towards generating meaningful metadata for digital music.

So, in essence, the motivation of this thesis arises from three different areas of research efforts. Firstly, derivation of semantic interpretation from acoustic properties of music (Whitman, 2004) only considers simple mapping of acoustic metadata to consumers' tags. Such simple mapping is not suitable for semantic search because it provides no structured mapping of machine level statistics. Secondly, the Motion Picture Expert Group (MPEG) initiative to create a set of standard syntactic vocabulary for multimedia content enabling search and retrieval of the content is known as the MPEG-7 standard. The MPEG-7 Audio part defines syntactic metadata to represent acoustic properties of the audio content. Music audio described using MPEG-7 audio descriptors could lead to automatic search and retrieval of the music materials. But, MPEG-7 audio descriptors based acoustic metadata is understood neither by music producers nor by consumers i.e. ordinary listeners. Thirdly, semantic web initiative provides standard tools and techniques to represent contents in way that are both human understandable and machine process-able but do not stipulate how to annotate/categorize music using structured metadata.

This research will utilize standard acoustic feature based audio description schemes (MPEG-7 Audio) to fulfil the purpose of music annotation by music producers to enable effective retrieval of musical resources by music consumers in general.

1.2. Contribution

The main contribution of this thesis comprises the development of a structured semantic vocabulary in order to annotate digital music; based on the idea that if music items can be annotated semantically then search and retrieval may be enhanced. Given the state of the three research fields mentioned earlier, this research will address the following three issues:

1. The MPEG-7 audio based standardized acoustic metadata generated by automated tools (e.g. Crysandt, 2005) are not suitable for annotation by ordinary music users (both producers and consumers).
2. Perceptual Metadata generation attempt (Whitman and Lawrence, 2002) tries to derive perceptual metadata from acoustic data and leads to generation of unstructured syntactic tags.
3. Semantic Metadata vocabulary for music takes at least two different prominent paths. One (Raimond et al., 2007) develops structured semantic metadata for use by music domain only without any mapping to the acoustic metadata; whereas the

other (Hunter, 2003) develops upper level multimedia (audio) ontology (semantic multimedia metadata). But none of these provide any means how they can be utilized for music annotation by ordinary music users.

Currently, the use of standardized meaningful metadata is becoming a dominant industry wide trend because of the frequent use of both syntactic and semantic metadata to improve search and retrieval efficiency. In the case of musical objects, the figure 1.1 shows that generation of metadata has taken two disjoint directions: some metadata collections are purely syntactic (pink block) while others are fully focused on incorporating meaning or semantics (green block) into it. Acoustic modelling is used to generate acoustic metadata from musical objects that represent features of music audio extracted and analysed using digital signal processing techniques. Further analysis has also been applied to acoustic metadata by using perceptual modelling to generate perceptual metadata that actually categorize musical objects using a simple set of keywords. Perceptual metadata presents only a set of syntactic metadata that are used to group musical objects based on their acoustic metadata and values without proposing any semantic interpretation.

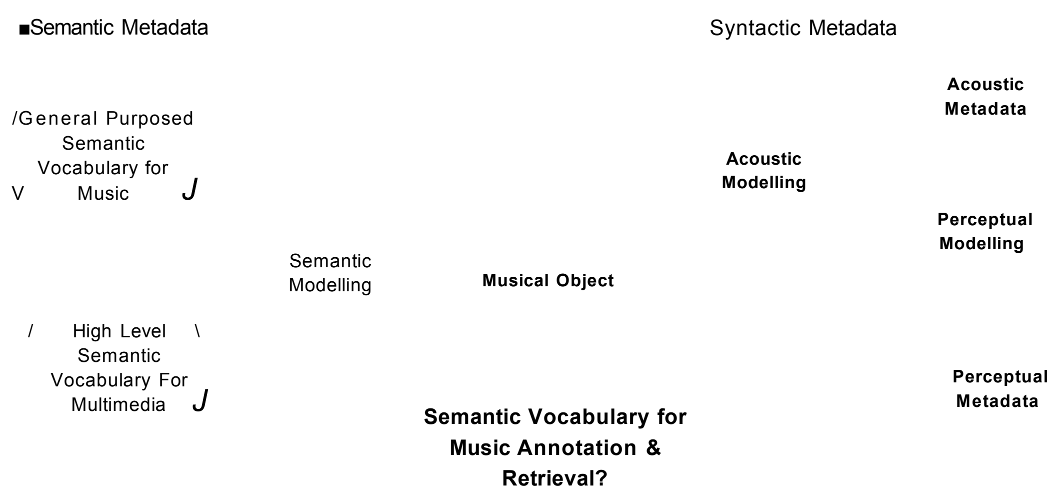


Figure1.1: Metadata Modelling for Musical Objects

Semantic modelling of music has taken two distinct directions. First one establishes a structured set of metadata that is organized with meaningful representation but is aimed to be used for general purpose editorial metadata (e.g.) for music. The other one considers defining semantic vocabulary for multimedia objects only. Such high-level semantic vocabulary describes music only as a specialization from Multimedia segment – it but does take into account the specific details of the representation of musical objects/music segments as understood and utilized by ordinary music users. The figure 1.1 highlights this gap using orange rectangle. There are initiatives such as (Raimond et al., 2007) that aim to develop meaningful music metadata (namely a Music Ontology) that focuses on structured and meaningful representation of music metadata but does not provide any relation of this metadata with acoustic/perceptual features so that effective retrieval can be achieved while searching.

Acoustic modelling of musical objects has created acoustic statistics based syntactic metadata and perceptual modelling of acoustic metadata has generated another type of syntactic metadata called perceptual metadata. Perceptual metadata in most cases does not conform to any standard tagging scheme but there are two kinds acoustic metadata collections depending on whether they conform to standard tagging scheme or not.

Semantic modelling of musical objects is primarily inspired by the concept of incorporating machine & human understandable meaning. Some pay more attention on the representation of musical artefact related data for different levels of music users by defining general purpose semantic vocabulary for music; while others, pay more emphasis on the conformance to industry wide content representation standard and thus create high level semantic vocabulary for music media.

Perceptual meaning modelling from acoustical representation of music has been done by using automatic machine learning techniques utilizing community specific metadata (Whitman2002); however, it is not suitable for efficient search and retrieval of music in general. Because, these sorts of metadata do not posses enough structured & meaningful representation so that search algorithms can infer satisfactory results from a query.

Acoustic/signal level description of audio has been done by (Crysandt 2005). But, these audio statistics based techniques do not stipulate any rules to project audio signal features to different levels of music information that may be utilized for music categorization that will actually be able to link audio features with semantic contextual features e.g. whether a music is funky/loud/ soft or romantic.

Current efforts to develop music metadata poses an open question on how to develop semantic vocabulary for music annotation and retrieval; but at the same time making it machine and human understandable as well as conforming to music content representation standard. In the previous section, it was argued that use of acoustic metadata could be used for effective search and retrieval of music if meaning of those metadata could be made to be understood by the stakeholders. Our proposal here is to utilize the acoustic/perceptual music metadata in the generation of a semantic vocabulary for digital music.

The goal of this work is to enable music producers to categorize/tag music resources with meaningful metadata that will in turn contribute to effective search and retrieval of music. To accomplish that goal this thesis aims to create a set of semantic vocabulary that will enable music producers to annotate music media with keywords/concepts from the proposed set of vocabulary (that is understood by them) without being concerned about the underlying mapping and interpretation of acoustics or perceptual metadata.

Based on the key findings mentioned earlier and in the context of music search and annotation the following goals need to be addressed:

1. **Meaningful search results:** Delivering search results that reflect the meaning intended by the user (user query) needs to be addressed within the current capability of state of the art search techniques that relies mostly on keyword(s) based queries.

2. **Automation of Content based searching:** Associate implicit semantics with music media resources that will enhance content based searching requiring the maintenance of a balance between manual and full automatic annotation of the resources.
3. **Semantic Bridging:** Bridge the absence of connection (mapping) between keyword and content based tagging with semantic association.
4. **Enriching MPEG-7:** Enrich MPEG-7 Audio description with meaning that would yield useful response to search requests to overcome the limitations of providing normative descriptors and schemas to specify multimedia in general.
5. **Enhancing the ability of state of the art search engines:** Incorporate implicit semantics associated with audio material to provide the user to search through the content.
6. **Need for standardized vocabulary:** Develop a standard set of vocabulary that incorporates general (low-level) audio features with high level meaning intended by ordinary users during search.

Considering the goals mentioned above mentioned goals, two basic research questions may be synthesized:

1. How to annotate music against MPEG-7 description to deliver meaningful search results?
2. How to formulate a query against MPEG-7 description to satisfy search request?

This thesis specifically will aim to address the issues related to **question 1**, where the main contribution will attempt to reduce/narrow the absence of connection between keyword based tagging AND content based tagging using semi-automatic annotation of multimedia resources.

To achieve the goals we already have mentioned we will proceed as follows: At first, the requirements for music annotation by ordinary users will be identified, and secondly, the concept and properties of the music annotation vocabulary will be generated based on standard representation of musical audio content (Salembier 2007). This work will extend the Music Ontology by Raimond et al. (2007) based on the core multimedia vocabulary ontology by Hunter (2003). The research effort presented here is closest to the work of Hollink (2006) that created a visual ontology from existing standard metadata for annotating images. Finally, the use of the developed music annotation vocabulary will be demonstrated using an annotation tool developed (designed for the purpose of music annotation) to show its applicability with the identified requirements for music annotation.

1.3. Thesis Outline

The purpose of developing the music annotation vocabulary by modelling standard acoustic/perceptual features is to enable tagging/annotation of music to enable effective search and retrieval of music motivated by semantic search capability envisioned by semantic web initiative (Miller and Swick, 2003) and standard representation for music metadata. So, our literature survey presented in chapter 2 is divided into three sections. It starts by analyzing the state of the art of semantic search and retrieval in section 2.1 where the primary focus will be on two issues - firstly, it will attempt to identify the current/existing trend in semantic search technologies and secondly, it will try to understand several levels of representation of standard music metadata. Section 2.2 looks at the existing techniques and tools of semantic annotation in order to find the applicability and scope of usage of the proposed music annotation vocabulary. Section 2.3 presents a thorough study of the existing multimedia metadata standard by studying different descriptors and description schemes and specifically the audio description tools as well as already defined semantic multimedia vocabularies that are compatible with dominant standards.

The core contribution of this thesis has been detailed in chapter 3 which is divided into four sections - starting with the research methods followed to achieve and establish the contribution in section 3.1. Then it will touch the subject of different levels of information contained and conveyed by music to build the foundation of the proposed contribution in section 3.2. Section 3.3, establishes the essential underpinnings of the semantic annotation vocabulary and then the methods and principles of developing the proposed vocabulary have been detailed in section 3.4. The chapter 4 presents an evaluation of the developed vocabulary using different dimensions. Finally, section 5 concludes by revisiting the research objectives and presenting a critical analysis of this effort by performing a backward tracing of the achievements of this thesis with reference to the relevant research fields as well as directions for further improvement.

2- Literature Survey

The core contribution of this thesis aims to develop a set of standard and strucutred vocabulary to annotate musical objects with semantic metadata. To achieve that objective, literature survey under this thesis have been organized as the study of the state of the art discussing issues and prospects from semantic search and metadata, semantic annotation and standardized semantic metadata for annotation of music in three different sections. At the present state of the technology it is widely believed that the use of semantically annotated music information will ease the process of finding of musical objects. To determine the scope of semantic annotation as a facilitator for semantic search, the state of the art/practice on semantic search and retrieval in order to provide meaningful search results has been investigated in section 2.1 ; highlighting the relevant issues on semantic search. The existing semantic metadata, annotation tools, theoretical basis for conceptual representation of metadata, different initiatives to generate multimedia metadata as well as annotation methodology for utilizing semantic metadata will be examined in detail in section 2.2. Finally, section 2.3 of the literature survey takes a closer look at the standard multimedia metadata schemes that will later be utilized to create the proposed semantic vocabulary for annotating music as well as aspects of musical information that will be modelled.

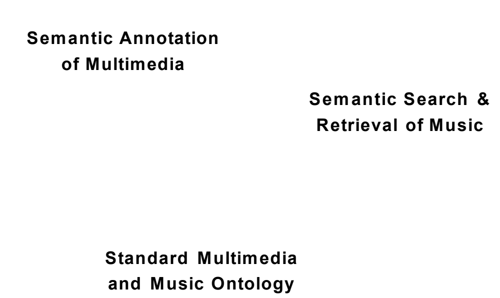


Figure2.1: Multimedia Metadata modelling

Figure 2.1 shows that the three underpinning key areas for the proposed semantic vocabulary for annotating musical objects and the intersection shown in orange colour represents the core contribution under this thesis by synthesizing these works in a novel manner. Our plan for literature survey includes three key areas –

- Semantic vocabulary for music annotation, search and retrieval
- Existing multimedia ontologies and
- MPEG-7 multimedia description features

The proposed vocabulary for annotation of digital music will evolve from these three broad topics as shown on figure 2.1 with orange circle. Effective semantic search and retrieval is largely dependent on two aspects – firstly, through the design and development of high performance and intelligent semantic search algorithms – which is limited by the fact that - search algorithms working on unstructured metadata need to capture diverse users' interest is a challenging problem to solve. As a result, fully automated retrieval of music through semantic search algorithms (without human intervention) returns unsatisfactory search results. Secondly, effectiveness of semantic search & retrieval may be achieved through structured representation of diverse users' metadata- semantic annotation of multimedia using structured metadata (ontology is an example of representing structured concepts) enables us to achieve effective results by search algorithms. When users' search terms are organized using semantic concepts from ontology then the task of search algorithms become simpler. So, in this thesis, I have tried to look at three broad areas – namely, semantic search and retrieval of music, semantic annotation of multimedia, and standard multimedia and specifically music ontology. The state of the art of semantic annotation of multimedia has been explored to determine how semantic annotation can lead to effective search & retrieval of music, and then existing techniques used for semantic annotation of multimedia have been explored so that similar techniques may be used/adapted to generate useful semantic metadata using the proposed music ontology concepts. The contribution of this thesis will present the design of a new music ontology that incorporates standard description of acoustic metadata.

2.1. Semantic Search and Retrieval of Music

2.1.1 The General Search Problem and Semantic Search

The rapid growth of the advancements in internet technologies and increasing collection of audio resources on the internet have made it possible for music listeners to access huge volume of music & songs online. But, music information retrieval needs to be tailored to fit the tastes and needs of individual listeners (Li and Ogihara 2006). Traditionally, music files are organized using song title, singer and other textual tags but such techniques are insufficient for modern users (Gang et al., 2008) to find their music items of choice. Due to increasing number of music collections traditional browsing of folder hierarchies or search by title and album name tend to be insufficient and as a result finding novel ways of music resource organization has become a research issue (Goussevskaia et al., 2008).

The **general search problem** of digital audio is a huge challenge due to the lack of contextual information available to guide the search (Koister 2009). The problem with traditional search engines is of lack of semantics or meanings; they can only find pages that contain the chosen key/search/content word in the text, as a result finding relevant information sometimes becomes impossible. Major search engine vendors as well as scientists predict the future of search engines lies in its semantic capability (Hai et al., 2008). But they all have different opinions on how the semantics (or meanings) should be incorporated on the search query as well as the content itself. Two basic views of a semantic search are found in the contemporary literature and these are identified by the location of the semantic resources to be implanted (Berkan, 2007) with the content of the resource.

Implanting
Semantic Search

Content
Categorization

Content Search

Figure2.2: Content Categorization Vs Content Search

Location of implanting semantics with the information content may be either applied during categorization/annotation or during searching the content as shown in figure 2.2. First view follows the idea of embedding semantic resources in the web pages themselves; and then search engine architecture will apply semantic reasoning and then create semantic index for semantic retrieval as well as including traditional keyword index based retrieval. This view of semantic search utilizes semantic annotation to associate resources with concepts or instances from structured domain knowledge (semantic vocabulary)and refers to the process that uses semantic annotation algorithm to associate concepts or instances from domain knowledge, annotates documents creating domain resource repository, and generates semantic index repository e.g.in (Zou et al.,2008). According to user's query keyword, search program performs a search task from a semantic index repository and the search results corresponding to the semantic features are returned to user. One way of executing search algorithm against such index repository is to expand the query keyword (if no instance set or direct descendent concept set exists) by utilizing a matching predecessor concept set. The big advantage of semantic expansion search is that a fewer number of possible search results are created (Lamberti et al., 2009). But, this kind of approach is limited by the assumption that every web author will have to abide by the complex rules of semantics and interoperability issues among standards. At present, there exist no standard compliant domain ontologies for annotating diverse types of musical objects that can utilize such semantic expansion search.

The other view is to locate the semantic resources in search engines which deploy algorithms to analyze the semantic information; examples of such implementations are

Powerset, Cognition, Lexxe etc. (Midwinter 2007). Google sees semantic search technology as part of the algorithmic mix, not as a replacement to its traditional keyword-analysis approach (Perez 2009). But, creating semantic resources e.g. the creating knowledge of languages on which search algorithms will work is still considered as an expensive endeavour. However, the task of content categorization becomes a distinct phase in the life of a content starting from content creation/publication phase to content consumption (search/retrieval) phase (refer to the figure 2.3).

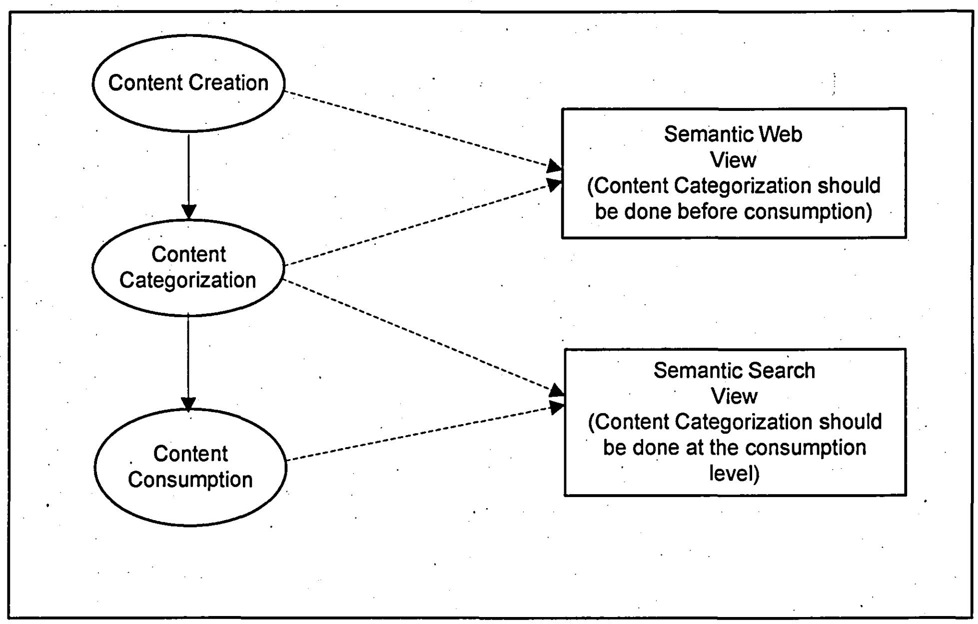


Figure 2.3: Views on Semantic Search

According to the first view, content categorization should be pushed more towards the creation/publication phase whereas the second view pushes the categorization completely towards the consumption stage. Apart from this, these two views have one thing in common; both of them enable semantic search and retrieval and rely on generating semantic knowledge associated with the resources content. Based on the above observation, it can be assumed that the idea of semantic search capabilities are to be used to complement and enhance search technologies by mapping the terms in user's search query to relevant semantic knowledge entities (Fernandez 2008).

At present search engine providers are using various algorithms to enhance search experience. Study of the state-of-the-art literature shows that the algorithms used by the search engines fall into two main categories based on the level of analysis it performs. Most algorithms used by search engines work with the keywords in the search query; though very recently, a new approach has emerged that apply sentence level query analysis (MacManus, 2007) to design search algorithms. But the direction of this is out of the scope of this research because this thesis aims to design a semantic structured vocabulary to annotate musical objects rather than designing a search algorithm.

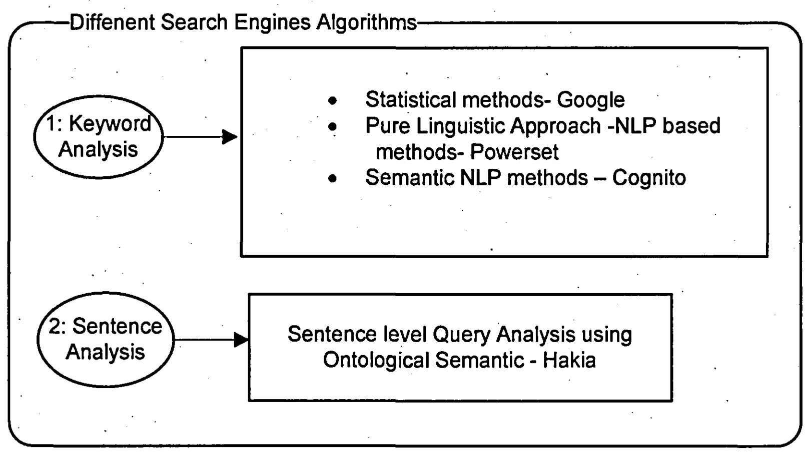


Figure 2.4: Classification of Search Engine Algorithm

Figure 2.4 illustrates two classes of search engine algorithms. Generally, keyword based search algorithms rely on finding the frequency and or relevance of the keywords that are present in the search query with the entire link structure of the web pages to determine which pages will be most important to list as a result to the search query. To do this Google (Google, 2008) conducts hypertext-matching analysis to determine which pages are relevant to the specific search being conducted, combines overall importance and query-specific relevance using statistical methods to put the most relevant and reliable results first. Powerset (Powerset, 2009) attempts to improve the user experience using Natural Language Processing (NLP) technology with an advanced parser focusing upon 'linguistic' knowledge. Basically, Powerset works on the search keywords using brute force computation and relies upon understanding the *structure* of language in preference to building a large database of terms and synonyms. Cognition (Cognition, 2008) uses the semantic mapping of the English language by combining semantic map and linguistic elements to optimize semantic understanding such as word morphology, grammatical structure, semantics (word and sentence meaning, augmented by synonymy and taxonomy), spelling etc. From the "context understanding" view point, Cognition claims to apply "Semantic NLP technology" (that "understands" word and phrase meanings) in contrast to Powerset's pure NLP. Hakia semantic technologies (Hakia, 2007) attempts to analyze the concept of a search query using sentence analysis instead of keyword analysis as is done by other major search engines, including Google. It analyzes search query at sentence level and relies upon ontological semantics to rank search results. Both, Cognition and Hakia have considered semantic knowledge (though represented differently) to apply during the processing of search query at different level of granularity (key word and sentence level).

It is clear, given the present state of the art, that search engines consider semantic search to be implemented involving the analysis of the textual query only, based on the assumption that users are unable to type much more than a simple keyword for searching contents. So search engines (specifically concerning multimedia) are faced with the challenge to derive meaningful outcome in the search results. This leads to the challenge of how to represent and index the multimedia content for efficient retrieval (Chang and Amanda, 2007).

In order to provide personalized and context aware access to content (mostly digital multimedia contents) collected from different heterogeneous disjoint sources requires an understanding of the content as well as users using them (Abels et al., 2005). Considering the current state of semantic search, the main focus of this thesis is to identify content data from music resources and create an understanding for the music listeners who search for them. Motivated by this goal, study of different methods and issues related to music search will be presented in the next two sections.

2.1.2 Existing Music Search Methods and Techniques

Having studied the general problems related to semantic search I will now pay attention to more specific aspects of music search methods and associated techniques. Searching for music is done by anyone from ordinary listeners to music experts. An untrained music listener not only just enjoys the music s/he is also aware of the change of key, repetitions and resolution. But, during search for a musical item the ordinary listeners need to describe music in the form of keywords. This implies more information to be provided to ordinary listeners about music than that available to them at present (Storr, 1997).

As many more music collections are made available online more often users wish to retrieve music from an audio collection given a query, representing a portion of the music that is either sung/hummed, played, or otherwise encoded using text (Suyoto et al., 2008). Forming the search query using free text or hummed audio (that is expected to retrieve the intended piece of music) leaves ordinary music fans with unsatisfactory search results. As a result, the huge proliferation of digital music in the form of audio resources on the web presents new challenges for search engines about how to incorporate search by musical content (Ruxanda et al., 2008) for consumers.

Beneficiaries of music search and retrieval may be grouped into three categories: first the industry bodies who are mainly focused towards recording, aggregating and commercial dissemination of music; second, the ordinary listeners who are interested about finding and using music in personalized ways and third, the experienced music producers, performers, composers, teachers, musicologists who are knowledgeable about music. Users belonging to industry bodies and music professionals play the role of music producers in publishing music contents; while the ordinary listeners mainly take the consumer role.

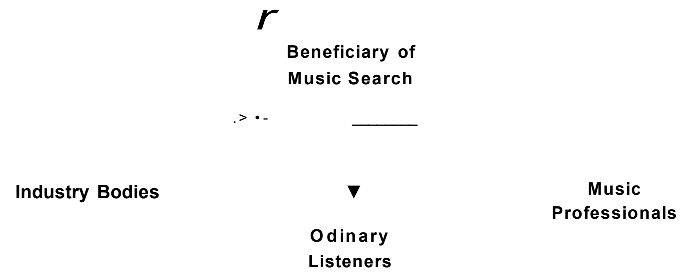


Figure 2.5: Different types of Music users

In the context of Music Information Retrieval (MIR) ordinary listeners (who are mainly interested about personalized search and retrieval of music) are one of the main beneficiaries (as shown in figure 2.5) of MIR systems and performance improvement of such a system depends on finding the method of music search that could lead to better understanding of how ordinary listeners (and/or music professionals) interpret music and what they expect from music searches (Casey et al.; 2008).

The proposed contribution (as detailed in section 1.2) will be targeted towards publishing of musical contents by music producers who can benefit from using the structured vocabulary to categorize the musical content. These categorized musical contents processed by search algorithms will in turn bring satisfactory results to the music consumers. In this context, I will now study the state of the art of general and semantic search methods used for music search, and how music tagging/annotation can improve music consumers' search experience identifying a set of requirements. The work carried out under this thesis will be targeted towards the annotation of music items by music producers so that the generated sematically annotated music information facilitates music consumers providing meaningful search results. So, the proposed music annotation vocabulary has been aimed to directly serve the purpose of music annotation by music producers; that will indirectly enhance the search experience of music consumers. In the rest of this thesis, the term 'music producers' will be used as the end users of the proposed contribution of this thesis.

Now, referring to the figure 2.6, methods to search music may be classified in to two broad categories based on the type of query allowed by the search engines: first one is the audio query that either contains sung/hummed audio segments or musical scores. The second one is the textual query that contains keywords.

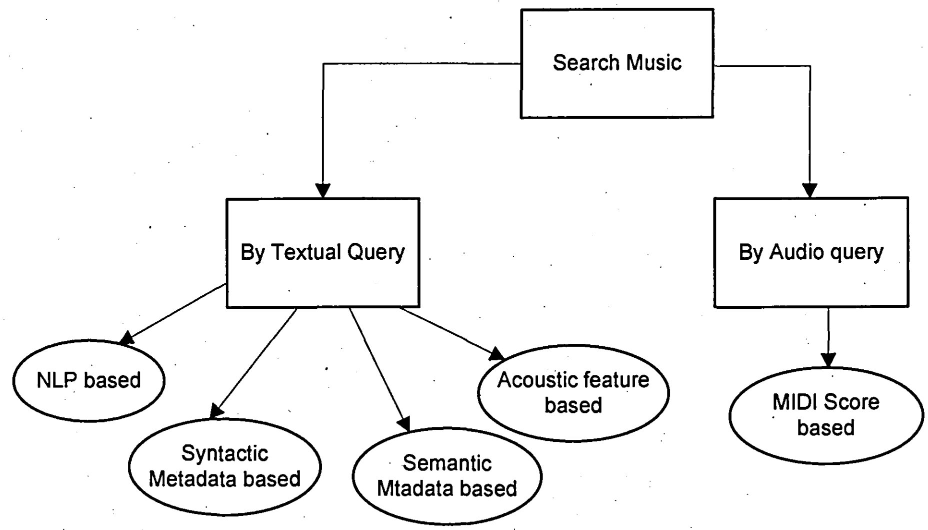


Figure 2.6: Different ways to search Music

Depending on how audio queries are expressed and analyzed, audio based search queries may further be classified as musical score based (MIDI score input) (Ghias et al., 1995) and musical metadata based (generated from humming input) (Maddage et al., 2006). But, such systems rely mostly on the end users ability to present query in the required format. Search methods based on textual queries may be of several kinds based on how they are being processed. Textual queries are processed usually on the basis of subjective metadata like title, album, artist/singer name, genre, style etc. Approaches used by major search engines and commercial music sellers fall under this type. Others attempt to enable processing search query using NLP techniques covering sound mood, cultural context, user profile and user defined metadata (Baumann et al., 2002; Celma et al., 2006; LastFm, 2008) etc. But, these systems suffer from several drawbacks such as how to generate the metadata (automatically or semi automatically) and filter noisy metadata as well as coverage of a large number of end users.

Automatic metadata generation from music resources has gained considerable attention in the contemporary research. For example, work presented by Knees et al. (2007) emphasizes the generation of metadata from web pages containing contextual information about music files - It combines NLP with semantic information related to audio as well as contextual metadata of audio links with low level acoustical features. On the other hand, effort by Kim et al. (2004) does rely solely on generating automatic metadata index from acoustical data. But, automatic metadata generation techniques require training data and often produce unsatisfactory tags.

Audio search and specifically searching for a song is inherently different than searching for any other multimedia resources on the web (Ali and Arabi, 2006). For example, search engines can display the top matches to users' search criteria almost immediately provided users keyword matches with song title, album or artist's name. However, in contrast, if the users' query is based on the concept presented by perceptual feature (e.g. timbre, melody or tempo) of the underlying sound or by musical property then how it can be ensured that this random part contains enough information for the search engine to locate the intended song by the user. Referring to the example presented in figure 2.7, I have tried[1] several music search engines to retrieve the famous song titled 'Lucky Star' by Madonna using the query characterizing the tune of the music as "lucky star by Madonna but faster"- this query contains keyword "faster" that is related to the tempo of the underlying music. But most of the music search engines available at present are unable to handle search query like "lucky star by Madonna but faster". Indeed they are unable to handle other keywords such as 'bright', 'sharp' describing tonal aspect of a musical piece; 'rising', 'falling' describing melody of the musical content.

It is evident that to enhance the user experience with semantic capability will require building structured semantics and knowledge that will create the foundation for semantic

[1] See Appendix B

search. Baumann et al. (2002) presented a search and query framework for human oriented Music Information Retrieval (MIR) system. The framework proposes to support for textual query and involves audio analysis to retrieve acoustic property and use natural language processing to analyse the query. To classify this framework depends on structured vocabulary e.g. ontological semantics. The proposed ontology in this thesis is aimed to support this type of textual query.

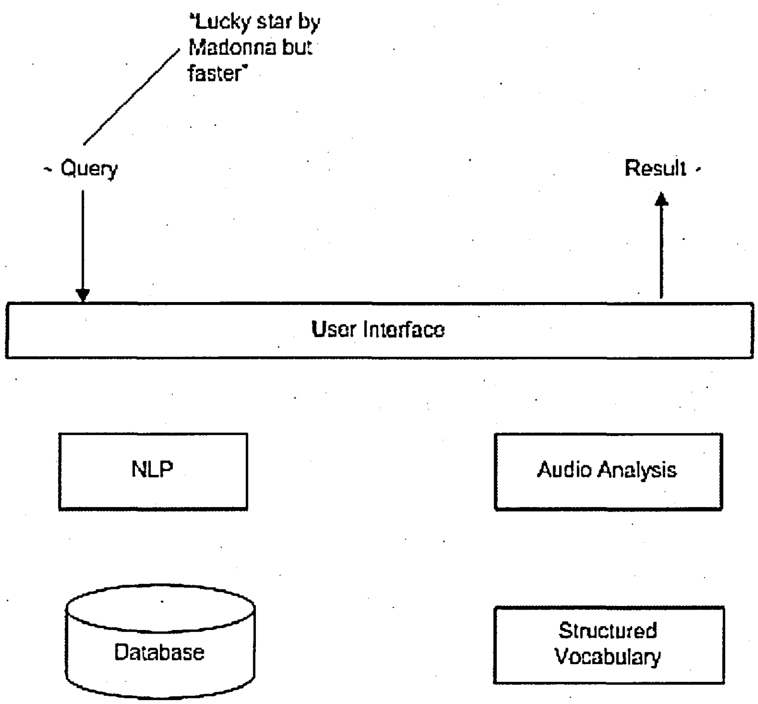


Figure 2.7: Search and Query Framework for MIR Systems –

adapted from Baumann et al. (2002)

As shown in figure 2.7 above, this thesis aims to contribute to this framework (Baumann et al., 2002) by designing a structured vocabulary. The presented framework considers textual query for searching music and uses NLP tool to handle phonetic misspellings and linguistic analysis of the keywords that are present in the query. Incorporation of semantic information as a structured vocabulary in the Search and Query framework is aimed to support integration of information about artist, genre, year, lyrics, as well as automatically extracted acoustic properties like loudness, tempo, and instrumentation by using semantic ontology as the basis for query processing to extract relations between these high level semantic concepts that will allow to satisfy semantic queries involving faster/slower, cheer-up/calm down, mix of calm-down and cheer-up music etc. The proposed vocabulary in this thesis will fall in the category of formal semantics as depicted below using the orange cylinder - the detail of which will be discussed in the actual contribution section. But for now let's focus on few key requirements that semantic search frameworks need to be fulfilled as sketched in figure 2.8.

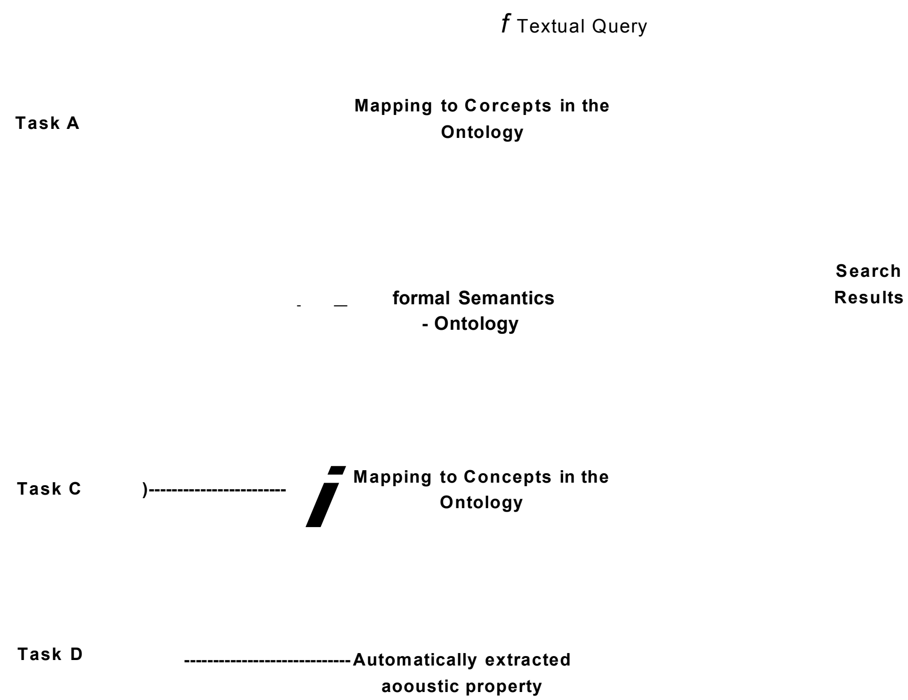


Figure 2.8: Requirements for Semantic Search and Retrieval of MIR Systems

At first, the choice of the semantic metadata must be made against which the text query will be answered (Task B). After deciding on the semantic metadata and structure the next task is to map the textual query to semantic concepts and relations (Task A). Then to relate the query to the content itself, automatic extraction of acoustic property requires us to decide what and how the features (standards and guidelines) of the musical content will be extracted (Task D). When the appropriate acoustical property has been extracted then these should be mapped to the semantic metadata structure to provide search results against query map (Task C).

Let's focus on how these requirements are being fulfilled. In the previous section, I reflected on several commercial initiatives by search engine vendors (Google, Hakia, Cognition etc.) on the mapping of textual query to semantic metadata structure that paves the way for solution to task A. In the arena of music metadata, there are research initiatives by Raimond et al. (2007)that creates a set of music vocabulary that may be leveraged with other taxonomies to express cultural metadata and music related content information e.g. by Rho et al. (2009) that extended music metadata concepts and relations to incorporate listener's mood, situation etc. Though these attempts to creating formal music metadata, shown as Task B (in figure above), can serve as a standard set of semantic metadata they do not provide how these metadata may be linked to the acoustic representation based content-based metadata of the music audio material.

Several significant works (Knees et al., 2007; Kim et al., 2004) that concentrate on extracting music's acoustical features but each has got different ways how do they use/extract those features (Task D). Knees et al. (2007) attempted to create a search engine for large music collections that searches on the music by assigning semantically related information to individual music pieces found on the web page that contains link to that music and the extracted text based information is complemented by audio-based similarity (both task D and C). Their method creates the understanding of how low-level audio features may be mapped to higher level semantic concepts but conforms to no standard semantics and taxonomy and it relies solely on the probability of finding information based on the web link. The work by Kim et al. (2004) presents audio classification based on low-level acoustic features conforming to standards like MPEG-7 (which will be elaborated on it in section 2.3) only without defining a further mapping to any formal semantics.

Whitman (2005) shows how low-level audio characteristics may be mapped to semantic concepts to understand meaning of several acoustic features but does not stipulate how these meaning may be utilized in a standard application and no conformance to standard semantics has been specified. All of these contributions are related to extraction of acoustic features and mapping to formal semantics (Task C and D) - address the tasks residing under the orange dotted line indicated on the figure 2.8. In contrast to those research efforts, this thesis intends to create a formal semantic vocabulary for annotation of music files that joins the gap between the two sides of the orange line by defining a Music Annotation Vocabulary that contain adequate concepts and properties to capture the acoustic features. The semantic annotation vocabulary will be created as an extension to Raimond's music semantics (Raimond et al., 2007) to support annotation of music files by music producers and the concepts and properties of the semantic annotation vocabulary will capture low level acoustic features providing a rich set of metadata to support tagging and annotation of music pieces that will in turn provide a better way to query for music contents.

There are several metadata schemas for music available[2]. Most focus on Common Western Music Notation representing particular writing system of music, music classification for commercial purposes but most of these contemporary efforts rely on syntactic tag matching techniques and suffers from limitations syntactic tagging. The next section will carry out a closer examination of existing tagging mechanisms and their limitations.

2.1.3 Music Tagging

Tags are generally, free text labels applied to musical content; typically applied by the publisher/producer or the consumer of the musical item. Usually tags are unstructured without any vocabulary limit. Tags are a channel for narrative and social interaction[3]. Social Tags (Turnball et al., 2008) are sets of individual tags; often known as Folksonomy generated from a user-created bottom-up categorical structure with an emergent thesaurus. Social tags' folksonomy (Weller, 2007) based representation contain weakly labelled unstructured free for all vocabulary. Such tags provide insights to user behaviour and language usage (e.g. rap/ hip-

2 http://www.recordare.com/default.asp

3 http://wikis.sun.com/displav/SocTagsMIR/Social+Tags+and+Music+Information+Retrieval

hop), grouping of items based on tags and users choice of tags, generating user profiles from tagging behaviour, track changes to tags on particular item over time, finding social groups with shared interests etc. Attributes of social tagging include tagging rights (owner, group, anyone), tagging support (blind tagging, suggestive tagging), tag aggregation model (e.g. set model), type of object being tagged (artists, tracks, albums, labels, playlists, clips etc.). Examples of social tags include tags from MusicBrainz, Amazon, MyStrands, iMeem, FreeSound, LastFM etc.

Tags are visible on music landing page, tag pages and have become visible almost everywhere. Users tag something using desktop software (Last.fm Player & Scrobbler), web site, add to library dialog, add tag button on item pages (like artist, album, track), every listed item (charts etc.) has multi-function button and flash player.

People use tags for various reasons (Lamere and Pampalk, 2008) - to build playlists by tagging tracks, categorise user profile and the music catalogue, and get recommendations based on tags or for future search and discovery. Tagging brings benefits to everyone though it is done mostly as a personal activity of fun. According to a recent survey (Weinberger, 2007), 28% of Internet users have tagged or content and everyday 7% of internet users tag or categorize content. Among the internet users the percentage of ordinary music listeners who use tagging is tend to be lower. So, this thesis targets music producers (instead of the ordinary music listeners) to be the primary user of the proposed semantic vocabulary.

Comparison of different tag collection approaches by Turnall et al., (2008) presents pros and cons of five music tag collection approaches. Among them surveys, social tags, game-based tagging basically does rely on human participation, and as a result require expensive human labour. The other two approaches - text mining and auto tagging rely on automatic methods requiring less human involvement but suffer for the need of computationally intensive training data. Songs that are not annotated cannot be retrieved and result in what is termed as the cold start problem. Such problem is caused by ***popularity bias*** in the popular songs (in the ***short-head)*** tend to be annotated more thoroughly than unpopular songs (in the ***long-tail).*** In relation to these tagging approaches mentioned above let's now examine their shortcomings and factors that cause these limitations in the next few paragraphs.

Cold start and obscure content: A newly released musical track suffers from the cold start problem as the user doesn't know that it exists. Often a new user intending to tag a musical piece of choice does not care enough to bother about long tail tags and creates obscure content which poses low recall of that musical piece for search &retrieval. Solution to such problem may be achieved from alternative sources of tags, autotagging, tag games etc.

Precision and Recall: Weak labelling gives rise to the low rate of precision and recall. Sparsely tagged items are most affected by weak labelling.

Ambiguity : Sparsely tagged items lead to ambiguity in social tagging caused due to polysemy and synonymy.

Tagger bias: Sources of social tags and tagging communities are not unbiased and they represent only sample of music listeners worldwide. They represent on average a relatively young audience, tech-savvy people and some regions are underrepresented (e.g. Africa, Asia) as well as some styles of music are represented stronger than others; e.g. classical music vs. indie/alternative. Some styles of music are covered better than others.

So, parse or inadequate tags result in cold start, obscure content, ambiguity as well as leading to low or no covering of certain style of songs. As a result, people interested in those less represented music styles will not be able to contribute equally with compared to those are heavily represented. There also exists game based music tagging approaches with purposely built games (Ahn and Dabbish, 2004); e.g. Tag a Tune, Major Miner, The Listen Game, MoodSwings, Heard It, BBC prototype Moose6 (Mahmoud, 2005). These games are designed from the idea that humans are best at solving some types of problems as they like to play games that are fun to play and in effect some hard problems can be solved. With purpose built games users may be able to tag at the phrase/clip level, solve problem of obscure tags resulting in the potentially high tagging rates. But, gaming approach needs to overcome not only the challenge of make addictive/fun game but also faces issues of superficial tags applied by non-fans and collecting weak labels. Another approach to collect tags could be hiring experts for survey or to hand label content – that could result in consistent, strong labelling with fixed structured vocabulary. But such human-labour intensive approach does suffer from the small pre-determined vocabulary and doesn't scale to the long-tail tagging. Besides it is very difficult to construct widely accepted taxonomy.

Web mining using Google could be among other sources of tags (Whitman and Lawrence, 2002). In Google using "<artist name>" for music review does retrieve text of top ranked pages as we see for text information retrieval techniques. But artist names are not unique identifiers. One solution to better search result could be adding information to query (e.g. album names) but then the cold-start problem arises again as it needs at least one web-page. Besides quality of "tags" is another issue. Using Web mined tags e.g. Neptune (Knees et al., 2006), Music Rainbow (Palmpak et al., 2006), MusicSun (Palmpak and Goto, 2007) could bring possible solution to the tag collection except it is unable to provide wide covering for different styles and newly released music items.

Autotagging is another idea that uses content analysis to automatically apply tags acquired from other sources (social tags, games, web crawling) can be 'learned'. New music or unpopular music can be autotagged with the 'learned' tags and can scale to the long tail. Generally, auto tagging systems relies on audio feature extraction modules and then matches the extracted features with pre-labelled example data to categorize the input item. In essence these systems creates standard model of classification from already annotated items.

Examples of such systems include MajorMiner[4], LabROSA (Mandel and Ellis, 2008), BRAMS (Bertin-Mahiex et al., 2008), CAL-UCSD (Turnball et al., 2007a) etc. These systems fail to utilize the different levels of content information that is encapsulated in the musical object as showed in figure 3.1. So, this thesis chose to enhance current approaches of music tagging by creating a knowledge based representation of musical content. The proposed structured annotation vocabulary will contain implicit association with low level acoustic properties of musical sound so that ordinary users will be able to tag the musical object without being aware of complex scientific representation of low level features. Moreover, low level features are easy to extract using automated tools.

2.1.4 Summary

The present limitations of music tagging and annotation that have been discussed in the previous section leads to several challenges faced by music discovery, search and retrieval systems. A sample query such as in Turnball et al. (2007b) -"This is **soft rock, jazz** song that is **mellow** and **sad**. It features **piano, synthesizer, ambient sounds**, and **monotone, breathy vocals.** It is a song with a **slow tempo** and with **low energy** that you might like to listen to while **studying**" is very challenging to satisfy using traditional search algorithms. Finding songs with queries like this will require us to find alternative solution for cold start problem and unstructured music vocabulary that could be solved partially (if not completely) by bringing together the advantage of auto-tagging and creating structured vocabulary from social tags.

Referring to figure 2.8 above, let's take a closer look at task C and then task B. Task C prompts us to create associations of acoustic property of musical resources to some formal semantics. The common form of creating association in the context of digital multimedia (in this context it is digital music) is often referred to as annotation (Ruvane, 2006).There are

---

[4] http://majorminer.com/

different ways of creating annotation (Fu et al., 2005). A closer look at the textual document reveals that four different types of association may be created. Firstly, is by the way of association building that uses links or relations by making notes and drawing symbols. Secondly, by creating annotations at the 'collection' or composite level, this refers to many subparts of a single text document. The third type consists of node-to-annotation links – where annotations that do not visibly refer to any particular document element, but are localized within a document part (e.g., a longish note written orthogonally to the printing on a page). Fourthly, 'standard' hypertext associations are used to create mapping from anchored portions of the text to notes or commentaries. Such a note might be written in the margin, the top or the bottom of the page, or on a separate piece of paper.

Apart from methods of creating annotations, depending on the level and granularity, annotations might be of two types. Content annotation considers using (e.g.) text selection, emphasis and adding notes and Structural annotations may be achieved by (e.g.) creating a logical structure that is different from the physical structure or linking from one place to another place in a book. Similarly, this work intended to establish a logical relationship among different levels of music information (as indicated later in figure 3.1) by creating annotations based on the main contribution of this thesis– which is a structured semantics for annotating music.

For digital photo/image, annotation is a process of labelling the semantic content of photos (or objects in photos) with a set of keywords or semantic information (Shu and Bederson 2007). This thesis developed the proposed (semantic vocabulary) ontology to contain individual textual instances so that users will be enabled to associate free text labels/keywords from the ontology instances that are related with semantic concepts/properties.

Annotated information creates association between keywords and the contents of photos in different ways such as direct importing (e.g. using filename of the image); image analysis (e.g. specifying low-level visual features such as colors, textures etc.); extraction from image context (e.g. textual description about the context of image when the image is embedded on a web page); and manual annotation of images by users explicitly deciding which information should be added to the photos etc. Besides, semi-automatic annotation approaches incorporates users' feedback into metadata that was automatically extracted. Again, automation of semantic annotation is still an open issue (Tsinaraki, 2007) as manual annotation of every single multimedia resource is cumbersome while complete automation of resource annotation suffers from the inability to interpret high level meaning.

In the context of digital music, both content and structural annotations take a different view than that of text documents and images. Music unfolds in time and information conveyed by different musical dimensions is quite distinct from text and images. For example, creating annotations at the composite level refers to many audio segments of a single music file partitioned by time dimension.

In relation to figure 2.8, task B requires us to model formal semantics (e.g. ontology) to generate annotations. Use of annotation ontology by extending features needed for annotation is not new (Dedek et al., 2008). In a similar way, this research aimed to develop an ontology for annotating music audio resources that will consider MPEG-7 descriptions of the corresponding resource and enable semantic description that can be further utilized for retrieving meaningful search result. In a novel way, this research proposed a general approach for annotating music files that considers both MPEG-7 objective data about the music audio as well as extends multimedia ontology for the purpose of annotating music files in particular. Most of the related works in creating multimedia ontologies have only attempted to create upper level multimedia ontologies and only few of them have focussed on creating customized vocabulary for a specific category (e.g. Greek music (Tsekeridou et al., 2006)) of music files.

None of them addressed the problem faced by music producers while annotating music items that they want to publish in general.

2.2. Semantic Annotation of Multimedia

2.2.1 Introduction

Annotation means adding information to the existing content or resources without changing the original. These annotations are meant to be share-able over diverse network, domain and/or users. Annotations are additional data/information that is tied to the content/resource in question; that represent information about the resources and arises from the interaction between the resource and its user. These characteristics serve to distinguish annotations from the general category of "just additional data". The semantic web initiative by World Wide Web Consortium (W3C) has established standards to make resource content automatically process-able by machines as well as human-readable. Such capabilities would be valuable for sharing thoughts and knowledge. As a result the Semantic Web could support other functionality that would enhance annotation in various ways (Passin, 2004): Machine-understandable annotations will enhance intelligent search and retrieval and for that it was required to pay attention to widely accepted knowledge representation techniques to create semantic metadata. There are many opportunities for improving the state of annotations, and Semantic Web technologies can improve the process of annotation. Currently available annotation systems have been explored to highlight some of their adaptability and weaknesses to meet the objectives of this thesis. Next, newly evolving systems for annotation have been examined to see how early Semantic Web technology is starting to work its way into this area. To identify the barriers/factors for achieving meaningful annotations that will contribute for effective retrieval, comparative study of the existing annotation frameworks and their suitability for annotating music items have been carried out in section 2.2.4. But, before going into further detail, next two sections will discuss the definition of semantic metadata and how they may be conceptually represented respectively.

The objectives for annotations should be considered much higher than a kind of simple metadata. Meta data is data about something else which is intrinsically associated with the subject, e.g. name of the author of a book or music album. Annotations, however, are generally about a third party's or users' thoughts, data, information, or experience. These things aren't intrinsically associated with the thing being annotated. Instead, they arise from the association between object (a book, a song or musical piece) and annotator (e.g. user annotating the object). Annotations can capture a user's experiences, thoughts, and feelings about the item being annotated so that they can be shared. There are multiple definitions of annotations in the related literature depending how annotations are created, how they are shared and utilized or as an enabler of machine readable meaningful metadata.

Both syntactic and semantic metadata are proven approaches of tagging resources to make them readable/ accessible to content management and search engines. Tim Berners-Lee's vision for the next stage Semantic Web requires not just tagging information with syntactic metadata, but annotating or 'enhancing' information with semantic metadata to enable machine understand-ability of the full context of what that information means.

Unlike syntactic metadata, semantic metadata are generated from completely different notion (Seth, 2007). Syntactic metadata describes non-contextual information about content, focussing on elements such as size, location or date of content/document creation providing little or no contextual understanding of what the document says or implies. This level of metadata is often the extent of many content management technologies. Semantic metadata are metadata that describe contextually relevant or domain-specific information about content based on a domain specific metadata model. For example, if the content is from the business domain, the relevant semantic metadata could be company name, symbol, industry, sector, executives, etc., whereas if the content is from the Intelligence/security domain, the relevant semantic metadata could be terrorist name, event, location, organization, etc. Metadata that offer greater depth and more insight 'about the document' fall under the semantic metadata category.

Another, requirement for semantic metadata is to provide fast, precise access to relevant resources across heterogeneous networks and between domains as well as resource discovery but cost of manual metadata generation sometimes undermine its prospect for efficient retrieval (Hunter, 2003). On the other hand, annotation offers further possibilities regarding the computation of the agreement between different annotators as well as the evaluation of a system against a certain annotation.

2.2.2 Overview of conceptual representation

Formal Semantic Representation:

There are various ways to represent semantics and concepts. Simplest is the controlled vocabulary that actually is a list of terms that have been enumerated explicitly. The more structured ones are taxonomy, thesauri and ontology respectively. Taxonomy is a collection of controlled vocabulary terms organized into a hierarchical structure where each term in taxonomy is in one or more parent-child relationships to other terms in the taxonomy. A thesaurus can be defined as "a controlled vocabulary that leverages synonymous, hierarchical, and associative relationships among terms to help users find the information they need." The word ontology is used to mean different things, e.g. glossaries & data dictionaries, thesauri & taxonomies, schemas & data models, and formal ontologies & inference. A formal ontology is a controlled vocabulary expressed in an ontology representation language. This language has a grammar for using vocabulary terms to express something meaningful within a specified domain of interest. The grammar contains formal constraints (e.g., specifies what it means to be a well-formed statement, assertion, query, etc.) on how terms in the ontology's controlled vocabulary can be used together. Taxonomies and thesauri may relate terms in a controlled vocabulary via parent-child and associative relationships, but do not contain explicit grammar rules to constrain how to use controlled vocabulary terms to express (model) something meaningful within a domain of interest (Pidcock, 2003). The term ontology has been applied in many different ways (Garshol, 2004), but the core meaning within computer science is a model for describing the world that consists of a set of types, properties, and relationship types. In taxonomy the means for subject description consist of essentially one relationship: the broader/narrower relationship used to build the hierarchy. The set of terms being described is of course open, but the language used to describe them is closed, since it consists only of a single relationship. Compared to taxonomy, thesauri could in theory be considered ontology where there is only one type, one property as well as context, and very few relationships. In practice thesauri are not considered ontologies because their descriptive power is far too weak, precisely because of this limited vocabulary.

Languages to encode Ontology: The idea of Semantic Web as envisioned by Tim Berners Lee (Berners-Lee, 2008) provides a knowledge infrastructure to explicitly represent conceptualizations of domain knowledge in the form of ontology in different levels (Ding et al., 2007).

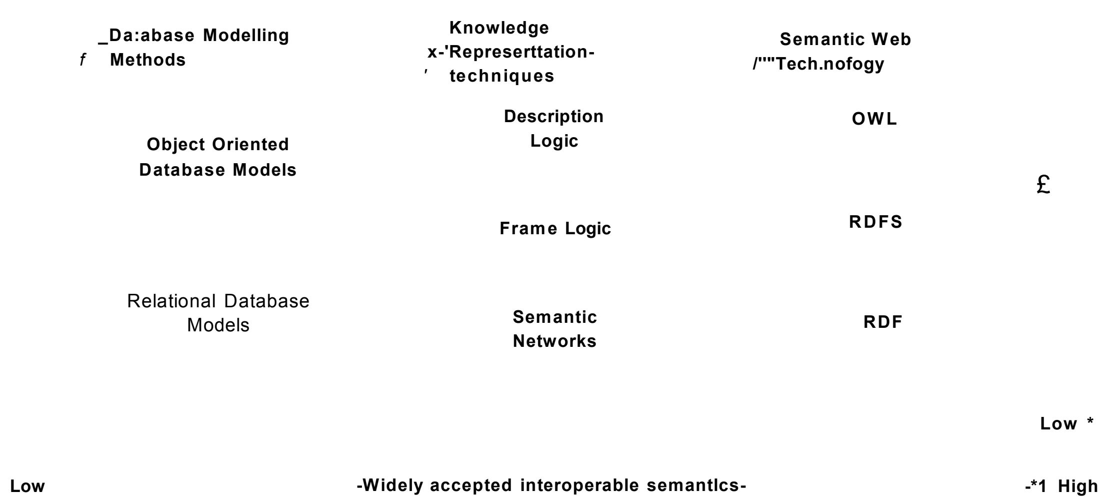


Figure 2.9: Evolution of Knowledge Representation Languages

The knowledge infrastructure inherited conceptual knowledge representation techniques from two other earlier paradigms such as Knowldge Representation (KR) formalisms, conceptual modelling methods for databases as shown in figure 2.9 [5].

Semantic web supports representation of ontology in standard languages namely the Resource Description Framework (RDF), RDF-Schema (RDFS) and Ontology Web Language (OWL). Referring to figure 2.9, the richness of the semantics represented in each paradigm increases from bottom towards top driven by the demand of porting implicit semantics into explicit representation. For example, *Semantic Networks* are characterized by their simple but powerful relational reference model in supporting conceptualization; *Frame Systems* incorporates additional constructs that model classes and instances in a user-friendly manner; *Description Logics* which came as descendents of Semantic Networks and Frame Sytems are highlighted by their formal semantics and decidable inference. Similar evolutions can be observed in the development of the databases. Common and accurate understanding of semantics across domains requires semantic descriptions to be interoperable. Database modelling methods are limited by their interoperability to be widely accepted across different domains. Besides, knowledge representation formalisms do not stipulate any common language format. As a result semantic web paradigm needs to devise common syntactic and semantics representation language which will provide common understanding across domains and between humans and machines (Wikipedia, 2008).

So, there are two basic issues that need to be addressed while encoding ontology. Firstly, the expressiveness of the underlying semantics presented by the ontology; secondly, the inference capability that can be achieved for the semantic representation in finite computing time. The more expressiveness the ontology presents the less the inference capability it poses. So, a balance needs to be maintained between the expressivity and the ability to inference.

---

[5] Figure 2.9 has been adapted from Ding et al. (2007).

There are several ways in which concepts of ontology might be represented (Stevens, 2001). There are a variety of languages which can be used for representation of conceptual models, with varying characteristics in terms of their expressiveness, ease of use and computational complexity. Languages currently used for specifying ontologies evolved from vocabularies defined using natural language to object-based knowledge representation languages such as frames, and then languages based on predicates expressed in logic such as Description Logics (Baader et al., 2007).

Vocabularies support the creation of purely hand-crafted ontologies with simple tree-like inheritance structures. Although this provides great flexibility, the lack of any structure in the representation can lead to difficulties with maintenance or preserving consistency, and there are usually no formally defined semantics. The single 'is-a' hierarchy type inheritance provided by a tree structure can also prove limiting while maintaining multiple inheritance hierarchies using single inheritance hierarchies is a difficult and arduous exercise. On the other hand *frame*-based systems are based around the notion of frames or classes which represent collections of instances; each frame may hold associated collection of slots or attributes which can be filled by values or other frames. In particular, frames can have a kind-of slot which allows the assertion of a frame taxonomy. But, the semantics of frame systems are not always clear about how to interpret an assertion that a slot is filled with a particular value. Complementary to frames is logic, notably *Description Logics* (DLs). DLs describe knowledge in terms of concepts and relations that are used to automatically derive classification taxonomies. A major characteristic of a DL is that concepts are defined in terms of descriptions using other roles and concepts supplying a number of reasoning services which allow the construction of classification hierarchies and the checking of consistency of these descriptions (Kifer et al., 1995). These reasoning services can then be made available to applications that wish to make use of the knowledge represented in the ontology. Frames generally provide quite a rich set of language constructs but impose very restrictive constraints on how they can be combined or used to define a class. They only support the definition of primitive concepts, and the kind of taxonomy must be hand-crafted. Description Logics have a more limited set of language constructs, but allow primitives to be combined to create defined concepts. The taxonomy for these defined concepts is automatically established by the logic reasoning system of the Description Logic (Baader and Nutt, 2002).

At present the most important ontology languages are Extensible Mark-up Language (**XML**), XML Schema (**XMLS**), Resource Description Framework (**RDF**), RDF-Schema (**RDFS**) and Web Ontology Language/ Ontology Web Langauge (**OWL**). Next, a brief overview will be presented of each of these languages (XML, XMLS, RDF, RDFS and OWL) based on the work by Antoniou and Hermelen (2004).

XML provides a universal surface syntax for structured documents but does not provide any means of talking about the semantics (meaning) of data. The same information may be organized in multiple ways with different order of tagging. So, there is no standard way of assigning meaning to tag nesting in XML. An example representation of XML description of an audio resource may look as this:

<Resource> <Audio> xyz </Audio> </Resource>

XML Schema is a language for restricting the structure of XML documents by only allowing defining new types by extending or restricting already existing ones. But, like XML, XML Schema also does not provide any way to assign meaning to tag nesting.

RDF is essentially a data model that provides building blocks like object-attribute-value triple, called a statement. RDF is domain independent and the users can define their own terminology in a schema language called RDF Schema (**RDFS**). Unlike XML Schema language RDFS defines the vocabulary used in RDF data model. Using RDFS it is possible create vocabulary that specify which properties to apply to which kinds of objects, value restrictions, relationships between objects etc. For example, we can write the following -

Audio is a subclass of Resource class.

This sentence means that all Audios are also Resource. So, there is an intended meaning associated with "is a subclass of". Thus, RDF/RDFS enables us to model particular domain through fixing the semantics of certain ingredients.

The fundamental concepts of RDF are resources, properties and statements. Resources may be thought of as any "object" or a "thing" e.g. people, location, event, book, author etc. Properties are a special kind of resources that describe relations between resources. For example, "written by", "contributed by", "contains", "shows", "features" and so on. Statements assert properties of resources. A statement is an object-attribute-value triple consisting of a resource, a property and a value. Values can either be resources or literals. Literals are atomic values (e.g. strings). The XML-based syntax of RDF is well-suited for machine processing but is not particularly human-friendly.

RDF and RDFS allow the representation of some ontological knowledge by providing modelling primitives such as subclass and sub-property relationships, domain (allowed type of values) and range (applicable classes for a property) restrictions and instances of classes. However, a number of other features are missing. Few of the limitations are: limitations of declaring range restrictions that apply specifically to some classes only. For example, using RDF/RDFS it is not possible to represent that "cows eat only plants". In RDF/RDFS only subclass relationships can be stated not disjointness of classes e.g. male and female are disjoint. RDF Schema does not allow boolean combination of classes (like union, intersection etc) e.g. we cannot define the class person to be the disjoint union of the classes male and female. Cardinality restrictions are also absent in RDFS e.g. "a person has exactly two parents".

OWL is an extension of RDFS in the sense that it adds extra language primitives to overcome the limitations of RDFS. To maintain the trade-off between efficient reasoning support and convenience of expression W3C's Web Ontology Working Group has defined OWL as three different sub languages - OWL Full, OWL DL and OWL Lite. OWL Full is the entire language which is fully upward compatible with RDF and added language primitives to RDFS to overcome the limitations that RDFS has and allows combination of primitives. But, its main disadvantage is lack of complete or efficient reasoning support. OWL DL is a sublanguage of OWL Full with some restrictions to ensure that the language corresponds to well studied description logic. Its main advantage is efficient reasoning support but it loses full compatibility with RDF meaning that every OWL DL document is a legal RDF document but not vice versa. OWL Lite excludes enumerated classes, disjoint statements and cardinality constraints. As a result, it is a further subset of OWL DL. OWL Lite gains in its easier implementation by users but lacks in expressivity. The choice among adopting the three sublanguages of OWL depends on the extent to which users require the expressive constructs and inference capability. **OWL** builds on RDF and RDFS and uses RDF's XML-based syntax.

For efficient modelling of ontology we need an ontology that has got richer express-ability than RDFS as well as inference capability. Generally, the richer express-ability is the more inefficient the reasoning support as well as computability. So, trade-off is required between express-ability and reasoning support when we choose an ontology representation language.

An extended version of OWL has been presented in Schneider and Horrocks (2006) – which is known as **OWL1.1** that actually extends OWL (specifically OWL-DL) by providing more expressive and computational power. OWL-DL lacks a number of expressive means. OWL 1.1 improves OWL-DL by creating an extension of logical basis of OWL-DL ontology language with disjoint, reflexive roles and negated role assertions added to it.

2.2.3 Ontology for Annotation

The widest knowledge expressed by natural languages cannot be manipulated by digital computers as this knowledge is not completely express-able in computable form. If a precisely formulated subset of the knowledge can be expressed using logic then it could be possible to process it for further inference by machines. According to Sowa (Sowa, 2000), the subject of knowledge representation provides theories and techniques from logic, ontology and computation to construct computable models of some domain where logic provides the formal structure and rules of inference, ontology defines the application domain and computation supports the implementation of logic and ontology using computer programs.

The logic is a simple language with few basic symbols. The level of detail of the represented knowledge comes from the choice of predicates depending on the domain. Logic needs to be interpreted with respect to predefined predicates for representing any subject to provide building blocks for defining the domain-dependent entities.

The definition of ontology in computer science does not seem to have settled down since early 1980s. It started with McCarthy's (McCarthy, 1980) interest of listing everything - "everything that exists, building ontology of our world". Then knowledge engineering (Stefik and Conway, 1982) was coined as an emerging discipline in computer science. Later, the initial interest of creating a catalogue of everything in the world turned into the efforts of applying the catalogue in solving practical problems as Sowa (1984) mentioned the Cyc Project (Lenat and Guha, 1990; Lenat, 1995), Electronic Dictionary Research (Yokoi, 1995), and WordNet (Miller, 1995; Fellbaum, 1998).

In the past ten years, the initial approach of cataloguing was being silently replaced by conceptual modelling of strictly task-oriented topics in a specific domain. The reason for such a shift from universal ontologies to domain specific ontologies is understood by scholars differently. Some considers the underlying factor for such a shift is due to engineering constraint (Sowa, 2006) while others consider it to be emanated from the inherent necessity for a knowledge engineering language (Gruber, 1993; Gruber, 1995) that comes as the definition of ontology in the form of the famous saying *"ontology is an explicit specification of conceptualization"*.

This definition of ontology seems to regulate the use of the term ontology in wide range of research and commercial projects - such as cataloguing, string matching, glossary of terms, thesauri or mentioning formal is-a instance or frames or in case of defining logical constraints etc. (Welty et al., 1999) though various research groups continue tailoring the meaning of the term 'ontology' so that it best fits and justifies their own research programs (Smith and Welty, 2001).

So, Guarino and Giaretta (1995) presented a revised version of the Gruber's definition as *"If we want to maintain its original (good) intuitions, we must weaken Gruber's definition, claiming that an ontology is only a partial account of a conceptualization"* and hence defines ontology as *"a partial specification of a conceptualization"* creating a quite a different understanding of the term *"conceptualization"* from quality perspective of the ontology.

Most ontological engineers accepts Gruber's definition that assumes ontology as a matter of knowledge representation with concepts conceived as human creations though there exists also different views that accepts Guarino's assumption of ontology as *"universal meaning expressed in formal logic"* and evaluates the quality of ontology in terms of reality representations (Smith, 2004). In (Shirky, 2005), an alternative definition of ontology has been presented from the classification and categorization perspective. Focussing on the Gruber's definition he argues that ontology is less concerned with what is than with what is possible to exist in a particular domain having relationship with each other as ontological classification or categorization relies on a set of entities into groups, based on their essences and possible relations emphasizing domain of the content and participants as the key factors in deciding where ontological categorization scheme will work well or not. The core attributes behind his view comes from the fact that meaning comes from human context not from the systems/machines. I also agree with this view that semantic concepts cannot automatically be associated with domain contents; knowledge representation in ontological format can present a way of aggregating semantic human context with physical domain information; where human participants will play the key role in creating association at least at the present state of the technology. The contribution in this thesis not necessarily aims at capturing the entire model of MPEG-7 semantics rather it will serve as light weight ontology to connect perceptual musical features with the acoustic content. Other than going into deep philosophical debate about ontology definitions and assertions the proposed light weight ontology is 'a model to compensate for the lack of content-based information while annotating music'.

Now, let's focus on the definition of annotation. An annotation has been loosely defined by the WorldWideWeb Consortium (W3C) Annotation Working Group (1995) as "any object that is associated with another object by some relationship". More specifically as mentioned in (Koivunen, 2005), annotations are "comments, notes, explanations or other types of external remarks that can be attached to any web document or a selected part of the document without actually needing to touch the document". According to the Oxford Advanced Learners Dictionary, annotation means: "to add notes to a book or text giving explanations or comments". These definitions can easily be extended to include multimedia resources and to allow other types of information than just explanations, comments and notes. Hollink (2006) has defined that an annotation as information that is explicitly related to an item with the purpose of describing the item for future reference and retrieval. However, none of these definitions include the formal semantic definition of annotation with respect to structured knowledge representation in the form of ontology. Therefore, this thesis defines an annotation as information that is explicitly related to a resource and such an annotation will formally represent the semantics of the metadata with reference to ontology; the task of semantic annotations is performed by tagging ontology class instance data and maps it into ontology classes (Reeve and Han, 2005).

2.2.4 Annotation Framework and Tools

Annotation frameworks are generalized platforms for creating annotations with few general requirements specified to support during implementation. Annotation tools are implemented using the general framework's requirements but during actual implementation such requirements may be implemented differently by different annotation tools. Annotation frameworks may be of different types based on their objective to create annotations for a particular type of object such as documents (limited types of format support), web services, multimedia resources etc. Other annotation frameworks are classified on the basis of format of annotations they create e.g. XML based or ontology based etc. As far as the type of annotation is concerned most of the annotation frameworks produce linguistic annotations.

Semantic Annotation Platforms (SAPs) are of three types based on the requirement of human intervention: manual, semi-automatic and automatic. Manually created annotations are error-prone and laborious task and lead to knowledge acquisition bottleneck (Suh and Bederson, 2007). Fully automatic creation of semantic annotations is an unsolved problem and instead, current systems focus on the semi-automatic creation of annotations (Reeve and Han, 2005). Semi-automatic means, as opposed to completely automatic, are required because it is not yet possible to automatically identify and classify all entities in source documents with complete accuracy. Strategically automatic annotation platforms can use pattern-based or machine learning based methods or may utilize methods from of these type. Pattern-based SAPs can perform pattern discovery or have patterns from manually defined rules. Machine learning-based SAPs utilize two methods: probability and induction. Probabilistic SAPs use statistical models to predict the locations of entities within text. Instead, semi-automatic creation of annotations annotation systems rely on human intervention at some point in the annotation process and consistently apply structured semantic vocabulary (e.g. ontology).

Generally semantic metadata may be created in two main ways. One is by identifying match words and meanings from dictionary/thesauri or by document analysis through application of rules and statistics for interesting patterns. Another widely accepted approach is the use of ontology as a method for creating semantic metadata effectively (Handschuh and Staab, 2003a). Semantic metadata paradigm based on ontologies suggests the use of ontology to guide the generation of metadata. This metadata is capable of facilitating the music consumers with content based discovery of resources irrespective of their locations and formats (Parekh et al., 2004). Advantages of utilizing ontology as annotation schema is many fold. The formalization of annotation schema as an ontology represented in standard language such as RDF/OWL meets the interoperability requirement of different conceptualization. Thus the use of standard encoding of annotation schema enables the reusability of the schema across different tools making the created annotations completely independent of the annotation tool actually used. Such RDF/OWL based ontological annotation model offers a general framework for the task of annotation that may be broadly applied to diverse contents. The fact that annotation is performed with respect to an ontological hierarchy offers annotators the possibility to choose the appropriate level of annotation detail.

Ontology based annotation tools use predefined concepts in ontology to mark-up a resource. The difference between regular syntactic metadata-based annotation and ontology–based semantic annotation is that in the former, the annotation is a plain text that is collected based on a fixed structure, while in the later, the annotation is a set of instances of classes and relations based on the domain ontology (Mostowfi et al., 2005). In ontology-based semantic annotation, the annotation process is the process of assigning the annotated object to a concept in the ontology (instantiating a class) or to a data type or relating it to another annotated text (instantiating a relation).

Through the annotation process meaning is associated to the multimedia resources (text, audio, videos, images), associating semantics to them through the attribution of one of the categories of ontology to each syntactic element of the representation language. The annotation of resources gives the opportunity to get a useful enrichment in terms of metadata. This process is accomplished by establishing relations among one or more elements that are present in a structure published on web (for example a web page) with a class of ontology. Due to this annotation process the knowledge can be shared not only with humans (for whom the existing web is designed) but also with software agents or machines.

The available surveys on annotation tools vary in the criteria they adopted to assess the tools. Uren et al. (2006) presented a survey of ontology based annotation frameworks and tools that helps us to categorize the tools based on standard formats and ontology support. But to achieve the research objectives of this thesis it requires to identify the tools based on the various content formats they allow annotating. Another survey done by Schroeter et al. (2006) focussed on content type (text, image, audio and video) the support but emphasizes the tool's ability to provide collaboration and sharing among different communities of users.

Experience from surveying tools independently, prompts to adopt a set of criterion to perform another survey that will better describe the research objective of this thesis and hence it focuses on existing annotation tools on the basis of following *three* requirements -

- Ontology based annotation for achieving interoperability
- Standard representation of annotations
- Type of content of heterogeneous multimedia resources

Table 1(a) & 1(b) provide an overview of different tools, systems and projects according to above mentioned three requirements. From table 1(a), it is clear that tools listed under simple syntactic annotation tools presented here fail to satisfy all three of these requirements. The rest of the tools as listed in table 1(b), though they all support ontology based annotations but were limited by their ability to provide annotations for diverse content formats (for music audio files). Most of them supported textual content (mostly HTML), very few of them could enable annotation for other types such as MPEG, JPEG, QuickTime, JMF etc. But, in practice there are a huge variety of digiral music media of different types such as audio (wav, mp3, ra etc.), and the musical object that comes as embedded in video resources(avi, mpeg etc.); for which none of these tools is able to provide generalized support.

Table 1(a): Survey of existing Simple Annotation Tools

	Annotation Tool	***Content format (Text/Image/Audio/ Video)***	***Standard representation of Annotation***
	Armadillo		RDF(S)
	KnowItAll		HTML
	SmartWeb		RDF, RDFS, OIL
	PANKOW		HTML
	WiCKOffice	Microsoft Office	Microsoft Smart Documents
Simple Syntactic Annotation Tools	AktivDoc	HTML	HTML, RDF
	Semantic Word	Word	DAML + OIL
	MagPie	HTML	HTML, OCML
	Lixto		Wrappers
	Mangrove	HTML	RDF
*	M-Onto-Mat-Annotizer	MPEG-7	XML, RDFS
	SemTag	***HTML***	***RDFS***
	Vannotea	JPEG2000/MPEG-2/Direct3D	XML
	Ontolog	QuickTime, JMF	RDF

Table 1(b): Survey of existing Ontology Guided Semantic Annotation Tools

	Annotation Tool	Content format (Text/Image/Audio/ Video)	Standard representation of Annotation
Ontology guided semantic annotation Tools	OntoMedia	QuickTime, JMF	RDF
	Amaya	HTML/XML	RDFS, XLink, XPointer
	OntoMat	HTML	DAML + OIL, OWL, SQL
	SHOE-knowledge Annotator	HTML	SHOE
	SMORE	HTML/Text/Image	RDFS
	Open Ontology Forge	HTML/Image	RDFS, XML, XLink
	COHSE Annotator	HTML	DAML+OIL
	MnM	HTML/Text	RDFS, DAML+OIL, OCML
	Melita	HTML/Text	RDFS, DAML+OIL
	Parmenides	HTML	XML (CAS)
	AeroSwarm	HTML	OWL
	KIM	HTML	RDFS, OWL
	Rainbow Project	HTML	RDF, WSDL/SOAP
	h-TechSight	HTML	DAML+OIL, RDF
	Thresher	HTML	RDF

So, the need for a generalized semantic annotation platform is evident. Another observation is that three of the tools (OntoMat, AeroSwarm and KIM) generate OWL formatted annotations while rest of them are confined to less powerful and older format of output annotation. Given this state of the semantic annotation tools, it is a necessity to design a semantic annotation tool that satisfies all three of the requirements mentioned above. To evaluate the proposed semantic annotation vocabulary for music media the proposed contribution will use a customized semantic annotation framework that will fulfil the proposed set of requirements for annotation tools and serve as an evaluation platform for semantic annotation as well as prove the applicability of the proposed vocabulary. The next section will present the detail study of the techniques used for conceptual representation in theory and discuss the rationale behind the choice for representing the proposed annotation vocabulary.

2.2.5 Vocabularies for Multimedia Metadata

Metadata is particularly useful for enabling computers and humans to efficiently access, organize and interpret data. The multimedia data is not self-describing and hence it is not possible for machines to interpret it. There are two types of multimedia metadata- ***content-based multimedio metadata*** derived from feature extraction tools is critical for describing multimedia data and another is ***semantic-level multimedia metadata*** to describe elements that the immutable multimedia data doesn't describe to allow for effective search and retrieval, content management, efficient access and delivery. For example, to determine whether a photo depicts a sunset, sports game, or particular person or event ***content-based metadata*** simply cannot meet the requirements for effective search and management of multimedia data.

According to Kosch et al. (2005) metadata's life cycle spans ***content***, ***metadata,*** and ***user*** spaces. In a content space, one can identify four major stages of content's life span - production/creation, postproduction processing, delivery, and consumption. Once multimedia content is created in the first stage, users can also modify it either by editing it or generating different versions of it. The metadata space involves two parts - metadata ***production*** and metadata ***consumption.*** The user space includes users that produce, process, and consume content. Content providers and producers are in charge of creating and producing content. They can enrich content by generating and attaching globally valid metadata. Processing users are those involved in postproduction processing and include those who index multimedia data in a multimedia database environment or those who prepare the multimedia content for adaptation, such as variation creation. Music consumers consume metadata and content. Their role is in browsing, searching, and consuming the multimedia data. They play an important role adapting as well as, for example, specifying their viewing preferences. Their roles can also be in completing the life cycle. For instance, a proxy server is an end user for the media storage server and at the same time a content provider of possibly modified content for the terminal consumers. So, end users who are mainly consumers may also act as processing users.

Multimedia metadata is also used for ***controlled terms*** such as taxonomies, controlled vocabularies and term lists, and classification schemes. For interoperable multimedia systems, it is necessary to use a standard set of type definitions (schemas) and standardized values in the description (Smith and Schirling, 2006). Generally, an interoperable description should depict at least following five discriminating characteristics6:

Representation: The primary (official) serialization format for the MM description standard, e.g. XML, RDF/OWL

Content Type: The type of media, a certain MM standard is capable to describe, e.g. Still image, audio, video, text.

Workflow: understood in terms of the ***Cononicol Processes of Media Production*** (Hardman, 2005) e.g. publish, process, creation etc.

Domain: The realm in which a MM standard is intended to be used in the industry, e.g. news, sports, music etc.

Industry: The (main) branch of productive (commercial) usage e.g. broadcast, music etc.

6http://www.w3.org/2005/Incubator/mmsem/wiki/Vocabularies

Table 2: Digital Content Representation Standards

Content Type	Existing Standards	Formal Representation	Available Mapping to RDF/OWL	Workflow	Domain	Industry
Still Image	Visual Resource Association (VRA)	VRA-RDF/OWL - non-XML	No-commonly accepted mapping	Publish	Culture	Archive
	Exchangeable Image File Format (EXiF)	EXiF-RDF/OWL non XML	Few available	Capture, Distribute	Generic	Digital Camera
	DIG35	DIG35-RDF/OWL	Available	Publish	Archives	Consumer
	PhotoRDF	RDF		Capture, Distribute	Personal media	Photo
Audio	ID3	non XML		Distribute	Generic	Music
	MusicBrainzMeta-DataInitiative 2.1	RDF		Production	Generic	Music
	MusicXML	XML		Production	Generic	Music
Still Image, Audio, Video	MPEG-7	XML, RDF/OWL	Available	Archive, Publish	Generic	Generic
	Advanced Authoring Format (AAF)	non-XML		Production	Content Creation	Broadcast
	MXF-DMS-1	non-XML		Production	Content Creation	Broadcast
	Synchronized Multimedia Integration Language (SMIL)	XML		Publish, Distribute, Presentation, Interaction	Generic	Web, Mobile Applications
	Scalable Vector Graphics (SVG)	XML		Publish, Presentation	Generic	Web, Mobile Applications
General Purpose	Dublin Core	XML, RDF		Publish	Generic	Generic
	TVAnytime	XML		Distribute	Electronic Program Guide (EPG)	Broadcast
Generic (Text, Still Image, Audio, Video)	MPEG-21	XML, non-XML		Annotate, Publish, Distribute	Generic	Generic
	XMP/IPTC	XML, RDF		Annotate, Publish, Distribute	Generic	Generic
	Flash	XML, non-XML		Publish	Distribution, Presentation & Interactivity	Broadcast
	MPEG-4 BIFS	XML, non-XML		Publish	Distribution, Presentation & Interactivity	Broadcast
	MPEG LASER	XML		Publish	Distribution, Presentation & Interactivity	Broadcast
	NewsML	XML		Publish	News	News Agencies
Video, Audio	EBU P/Meta	XML, non-XML		Publish	Generic	Broadcast

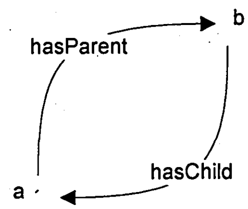
Based on these five characteristics for digital content representation standards, study of the existing standards for multimedia representation standards show that very little attention has been paid to support for various music content types and its usage by music producers to annotate and effective retrieval. Referring to table 2, there are lots of standards available to represent general purpose digital content and content type. In relation to audio materials most of the standards are concerned about publication and distribution of the content for commercial exploitation of the content. Very few of them did actually paid attention to musical objects and different levels of information content conveyed by those objects. None of them provided any further direction for enabling the music producers to facilitate annotation for further search and retrieval. Most of the existing audio content representation standards were designed to satisfy the requirements for specific domains for which they were created to document the workflow that support for a particular industry's purpose.

2.2.6 Summary

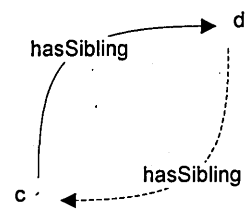
Annotation framework needs guidance from the ontology to allow sharing of knowledge and newly created annotations must be consistent with a community's ontology. If newly created annotations are created by instantiating arbitrary classes and properties the semantics of these properties remains void. So, the ontology is important in order to guide annotators towards creating relational metadata. Without the ontology it will be hard to establish more cues for assigning relationships between class instances (Handschuh et al., 2001). Most annotation tools make use of schema which specifies what can actually be annotated. These schemas can be understood as a formal representation of the conceptualization underlying the annotation task (Cimiano and Handschuh, 2003). As ontologies are formal specifications of a conceptualization (Gruber, 1993) it seems straightforward to formalize annotation schemes as ontologies.

Based on the intended methodology for annotating multimedia several tools and annotation frameworks have been developed and they fall in three general category. The first category is focused on the direct exploitation of signal level features for the purpose of annotation supporting a limited collection of content format (Schroeter et al., 2006) while the second one emphasizes the association of the resource as a whole with ontological concepts encoded in several standard representation and formats (Uren et al., 2006). Thirdly, among the attempts to create annotation frameworks and tools to join the above two, M-OntomatAnnotizer tool (developed under Acemedia project) (Petridis et al., 2006a) that enriches domain ontologies by using visual descriptor ontology (spatio-temporal ontology) represented in RDF. But the created ontology is limited to Image and visual feature descriptors only.

Advantages of utilizing ontology as an annotation schema are many folds (Kosch et al., 2005). The formalization of annotation schema as an ontology represented in standard language such as RDF/OWL meets the interoperability requirement of different conceptualization. Thus the use of standard encoding of annotation schema enables the reusability of the schema across different tools making the created annotations completely independent of the annotation tool actually used. Such RDF/OWL based ontological annotation model offers a general framework for the task annotation that may be broadly applied to diverse contents. The fact that annotation is performed with respect to an ontological hierarchy offers annotators the possibility to choose the appropriate level of annotation detail. In addition, annotation offers further possibilities regarding the computation of the agreement between different annotators as well as the evaluation of a system against a certain annotation. Moreover, ontology based semantic framework for annotation helps to constrain the possible relations between two concepts, thus reducing the amount of errors in the annotation process. Besides, description logic based ontological annotation allows inverse, symmetric and transitive and functional properties- thus representing an equivalence relation empower inference capability. For example, if entities a, b, c, d etc. are related using the properties like *hasParent, hasChild, hasAncestor, hasBirthMother* then following assumptions[7] may be made:

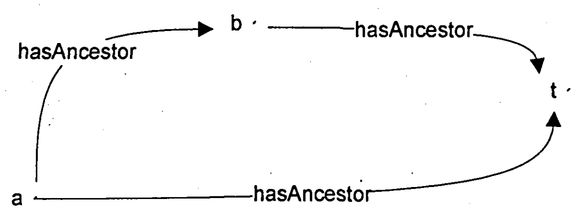


— Inverse Property



Symmetric Property

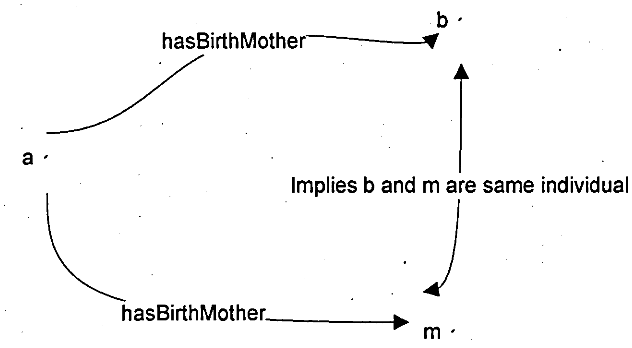
If individual entity **a** and **b** are related using *hasParent* and *hasChild* properties then these two properties are linking **a** to **b**; **b** to **a** respectively. So, *hasParent* and *hasChild* properties forms inverse property. The *hasSibling* property connects **c** and **d** vice versa; thus *hasSibling* is a symmetric property.



Transitive Property

If the entity **a** *hasAncestor* **b** and **b** *hasAncestor* **t;** then **t** is also an ancestor of **a** as shown above. Thus *has Ancestor* shows transitive property.

[7] Protege OWL tutorial



Functional
Property

If a functional property is defined between two individuals then there can be at most one individual that is related to the given individual. In the example above, **a** *hasBirthMother* **b**; **a** *hasBirthMother* **m**. Given that *hasBirthMother* is a functional property it implies that **b** and **m** are actually denoting the same individual.

2.3. Standard Multimedia and Ontology

According to the present state of the art, annotations of multimedia resources have been pursued in several directions each having different goals and objectives. Attempts to semantically annotate resources have mainly been focused on developing tools and vocabulary for annotating textual resources (Handschuh and Staab 2003b). Several efforts highlight annotation of images with ontology (Hollink et al., 2003) and specifically photographs (Wielemaker et al. 2001). In (Petridis et al., 2006b), an approach has been presented to enrich domain ontologies with low level video features from multimedia description standards (e.g. MPEG-7). To the best of my knowledge so far, MPEG-7 features from audio resources have not been used to annotate digital music to satisfy music producers need. To bridge this gap the proposed ontology will utilize audio features from dominant multimedia standard and will extend existing multimedia ontologies. The next section will introduce dominant multimedia standard for search and discovery of digital content and existing multimedia ontologies.

2.3.1. Overview of the MPEG-7 Standard

The Motion Picture Expert Group (MPEG) initiative to create a set of syntactic vocabulary for multimedia content enabling search and retrieval of the content is known as the MPEG-7 standard. It is also known as "Multimedia Content Description Interface" and it defines a standard for describing the **features of** multimedia content by providing a metadata system. The extensible MPEG-7 standard is subdivided into eight different parts; each part consists of different tools and elements to address different aspects of multimedia contents fulfilling different goals. The **Part 1 (Systems)** is intended for specifying the tools for preparing descriptions for efficient transport and storage, compressing descriptions, and allowing synchronization between content and descriptions. **Part 2 (Description Definition Language or in short DDL)** specifies the language for defining the standard set of description tools (Description Schemes, Descriptors, and Datatypes) and for defining new description tools. **Part**

3 (Visual) specifies the description tools pertaining to visual content and **Part 4 (Audio)** specifies the description tools pertaining to audio content only. **Part 5 (Multimedia Description Schemes)** specifies the generic description tools pertaining to multimedia including audio and visual content. **Part 6, 7 and 8** specifies guidelines & procedures for a) software implementation of the standard, b) procedures for testing conformance of implementations of the standard and c) extraction and use of descriptions respectively.

Part 3,4 and 5 of the MPEG-7 standard provide normative elements (Descriptors, Description Schemes) to describe low-level audio and visual features such as audio-energy, texture, color, motion etc with the expectation that most low-level features will be extracted automatically but producing high-level descriptors will require human intervention (Salembier, 2006). For extending the standard beyond the Description Schemes (provided by part 3, 4 and 5), the MPEG-7 standard has created provision of defining new Descripton Schemes (DSs) in the DDL defined under part 2, and to make those DSs available with the instantiated descriptions. So, part 2 of the standard details the creation of new DSs to avoid redundancy.

The part 5 of MPEG-7 standard, namely MPEG-7 Multimedia Description Schemes (MDS) has been designed to expand by combining individual descriptors and (or) other Description Schemes. So, Description schemes create provision for defining relationships between the constituent descriptors and description schemes. Description Schemes can be formed in two ways: 1) using specific audio or visual descriptors (e.g. low level features and signal structure, models and semantics) and 2) forming generic multimedia description (e.g. metadata related to the creation, production, usage and management). Description Schemes has been designed to expand by combining individual descriptors and (or) other description schemes. So, Description schemes create provision for defining relationships between the constituent descriptors and description schemes. The Description Definition Language (DDL) (as showed in figure 2.10) defines the Descriptors (D) and Description Schemes (DSs) by extending XML Schema language.

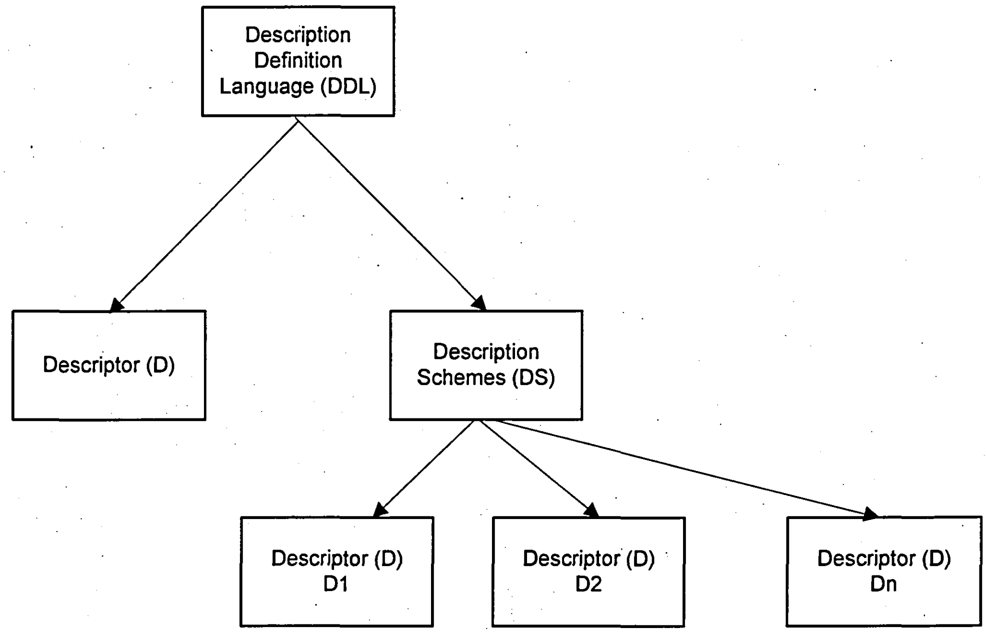


Figure 2.10: Descriptors and Description Schemes

The descriptors and MDS (as defined in Part 5) have been organized on the basis of their functionality based on the Basic elements. The functionalities considered are: Content Description, Content Management, Content Organization, Navigation & Acess and User interaction. Below, a brief overview of basic elements and content description functionality is provided as they are relevant to the main focus of this thesis. The **basic elements** are comprised of schema tools, basic datatypes, link & media localization tools and basic description tools. **Schema Tools** do facilitate creation and packaging of MPEG-7 Description. **Basic Data Types** are defined as a type hierarchy for base set of tools (description schemes, descriptors and header) from which all specific MPEG-7 tools are derived. The abstract types for descriptors and DSs are derived from Mpeg7BaseType.The audio (and visual) descriptors are extended from abstract DType and base type DSs are extended from DSType. **Link & Media Localization tools** are the basic elements and constructs for linking media files and localizing segments and regions. **Basic Description Tools** are the fundamental constructs and basic elements that provide specific data types and mathematical structures such as vectors and matrices. To provide **Content Description Functionality,** the MPEG-7 standard provides

DSs that includes two functionalities. One is **structural aspects** enabling the description of regions in an image, video frames or audio segments etc. The other one is the conceptual aspects in order to define real world semantics and conceptual notions.

2.3.2. Existing Multimedia Ontology

There exist different distinct types of multimedia ontologies in the recent efforts that created ontologies concerning description of image, audio and video. If these ontologies are categorized according to granularity, purpose, degree of formality & logical representations (Ceccaroni, 2001), it is found that LSCOM (Large scale concept ontology for multimedia) (Naphade et al., 2006) defines a general upper level concept lexicon, concept properties, and relations among concepts to describe multimedia. Extending this upper level general ontologies several domain ontologies have been created e.g. MediaMill 101 concepts are for broadcasted news video (Snoek et al., 2006) and pictorially enriched ontology for soccer videos (Bertini et al., 2007) are to name a few. There are other independent efforts that were initiated to fulfil the representational requirements for different domains e.g. representing medical education video (Luo and Fan 2006). Though these ontologies provide effective representation for the particular domain they were targeted for they lack in their applicabilty as a general upper level ontology suitable for representation in case of any multimedia content irrespective of domains. Moreover, concepts, properties and relations among concpets were not explicitly defined in most of these cases. But, formulation of properties and relations are very effective for successful annotation that leads to efficient search and retrieval. Depending on the purpose there are task/method ontologies, knowledge representation, linguistic ontologies. The Acemedia ontology (Petridis et al., 2006a) enriches domain ontologies using visual descriptor ontology represented in RDF. The visual descriptor ontology specifies spatio-temporal features for image representation only.

Modelling video pattern and connecting these pattern using linguistic relationship among concepts to associate them with video domain ontology have been proposed in (Bertini et al., 2005) for the purpose of semantic annotation task. The idea of using ontologies for knowledge modelling at the linguistic, perceptual, visual levels is still limited to particular domains. New domains will obviously require new linguistic descriptions, generalized abstraction to create open opportunity to model different domains. Bagdanov et al. (2007) proposes the use of several levels of multimedia ontologies to bridge the semantic gap between data and semantics where the multimedia ontology contains concepts at the abstract and linguistic level as well as the perceptual level. Perceptual manifestations in digital video are related with the linguistic facts through defined concepts.

Table 3: Comparison of Exisiting Multimedia Ontologies

Type	Ontology	Content Type	Workflow	Domain	Industry
Knowledge Representation Ontology	aceMedia Visual Descriptor Ontology	Video	Publish		
	Mindswap ImageRegion Ontology	Still Image	Publish		
	Visual Ontology forVideo Retrieval	Video	Publish		
Task Ontology	Common Music Ontology	Audio	Publish	Generic	Music
	Kanzaki	Audio	Publish	Generic	Music
	Music Production	Audio	Publish	Generic	Music
	Music Recommendation	Audio	Publish	Generic	Music

As presented in Table 3, the existing ontologies that covers music domain fall mainly on the category of task/method ontology to capture/provide definition of the relevant concepts & relations to specify reasoning process to achieve sharing, production or recommendations. All of these purposes serve for commercial exploitation rather than providing music consumers with satisfactory search results.

2.3.3. MPEG-7 Compliant Multimedia Ontology

Annotating audio and specifically music will require development of standard vocabulary that is enriched with low level audio features. At present there exist four prominent attempts to create MPEG-7 compliant ontologies. First one is the MDS Upper Laver represented in RDFS (Hunter and Little, 2007), which was later on revised to link to the ABC upper ontology and MDS was fully represented in OWL-DL8. Then DS-MIRF (Tsinaraki, 2007) model which is also known as MPEG-7 MDS ontology was proposed by Tsinaraki et al. (2004) that provides complete representation of MPEG-7 in OWL-DL. Another, the Rhizomik ontology model9 (Garcia and Celma, 2005) supports fully automatic translation of the whole standard MDS and Visual parts represented in OWL-DL. Finally emerges the COMM ontology model10 (Arndt et al., 2007) that re-engineers MPEG-7 using DOLCE design patterns.

Table 4: Existing MEG-7 Compliant Ontology

Ontology	Ontology Source	Description
ABC Ontology - MPEG-7 upper MDS ontology	http://metadata.net/mpe &Z	This is the MPEG-7 ontology was firstly developed in RDFS, and is now available in OWL-Full. The ontology covers the upper part of the Multimedia Description Scheme (MDS) part of the MPEG-7 standard. It comprises about 60 classes and 40 properties.
MPEG-7 MDS Ontology	http://elikonas.ced.tuc.gr /ontologies/av semantics ^zip	Based on Hunter 2001, this MPEG-7 ontology covers the full Multimedia Description Scheme (MDS) part (part 5) of the MPEG-7 standard. It contains 420 classes and 175 properties. This is OWL DL ontology.
MPEG-7 Ontology	http://rhizomik.net/ontologies/mpeg7ontos	Automatic generation of MPEG-7 ontology from MPEG-7 Standard with a generic mapping XSD2OWL has been implemented. The definitions of the XML Schema types and elements of the ISO standard have been converted into OWL definitions (Garcia et.al., 2005). The authors have also proposed to transform automatically the XML data (instances of MPEG-7) into RDF triples (instances of this top ontology). This ontology aims to cover the whole standard. It contains finally 2372 classes and 975 properties.

g http://metadata.net/mpeg7/

9 Rhizomik - MPEG-7 Multimedia Ontology, available from http://rhizomik.net/ontologies/mpeg7ontos

10 The COMM Ontology. Available online from http://multimedia.semanticweb.org/COMM/

The original approach by Hunter (2003) proposed a manual translation of MPEG-7 to RDF that was later translated to OWL-Full covering classes defining the media types (Audio, Audiovisual, Image, Multimedia, Video) and the decompositions from the MPEG-7 Multimedia Description Schemes (MDS) part (ISO/IEC 15938 –Part5). They started by defining RDF representation of MPEG-7 hierarchy of basic entities that are classified within MPEG-7 as Image, *Video*, *Audio*, *Audiovisual* and *Multimedia*; each having their own segment subclasses. A number of specialized subclasses are derived from the generic Segment Description Scheme that describe the specific types of multimedia segments, such as video segments, moving regions, still regions and mosaics, which result from spatial, temporal and spatiotemporal segmentation of the different multimedia content types. Also they presented RDF representation of basic non-multimedia entities within MPEG-7 as shown in figure 2.11.

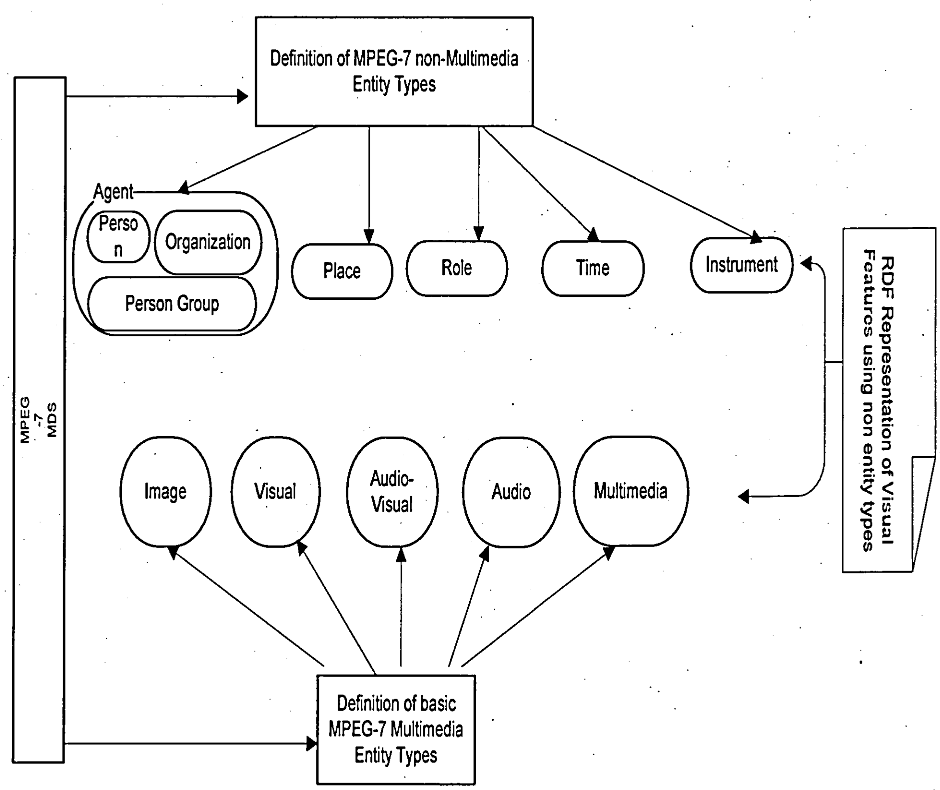


Figure 2.11: Upper level concepts from ABC Ontology

They chose to create RDF description of MPEG-7 MDS in connection with the visual features only. This was the first initiative to demonstrate the representation of MPEG-7 descriptors and DSs using RDF but suffers from the limitation of RDF expressiveness (specifically, range constraint, cardinality etc.) as well as interoperability of the RDF representation with other domain ontologies. This ontology has been used to describe the decomposition of images and their visual descriptors to enables queries for abstract concepts such as subclasses of events or agents to return media objects or segments of media objects.

Apart from ABC ontology's approach (Hunter, 2003) of representing MPEG-7 metadata in ontological form, Tsinaraki et al. (2004) attempted to define an OWL Upper Ontology, which fully captures the MPEG-7 MDS and thus forms a basis for interoperability between OWL and the MPEG-7 MDS. Based on this upper ontology they define the methodology for the definition of domain ontologies that considers the definition of domain-specific entity types to be represented by OWL classes that are subclasses of the appropriate Upper Ontology classes e.g. in a football tournament application the (domain concept) "FootballTeam" should be defined as a subclass of the "OrganizationType" which is an Upper Ontology class concept. Attributes that are not present in the super class are represented as appropriate object or datatype properties and additional constraints may be applied on the attributes inherited from the parent class. Relationships with additional restrictions compared with the ones of the general relationships defined in the Upper Ontology are usually needed then appropriate subclasses of "RelationBaseType" or of its appropriate subclass are defined and all the restrictions needed are applied to the newly defined classes. Such ontology is intended to support queries like "Give me the segments where Ronaldo appears or Give me the segments the shows events in Old Trafford" etc. Though this ontology attempts to solve the interoperability issues but will require determining how to index, annotate or filter the search results for each category of entity types.

Rhizomik approach (Garcia and Celma, 2005) is based on mapping XML Schema to OWL constructs that follows a generic XML Schema to OWL together with an XML to RDF conversion. Their contribution benefits the huge amount of metadata that has been already produced by the XML community by providing a translation mechanism to lift them for further utilization under semantic web framework.

COMM (Core Ontology of MultiMedia) (Arndt et al., 2007) is the first attempt to address the issue of semantic annotation of multimedia. It has been designed manually from MPEG-7 intended semantics and is based on DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering).The Description and Situation (D&S) and Ontology of Information Objects (010) patterns from DOLCE are extended into various multimedia patterns that formalize the MPEG-7 concepts. Digital data pattern is specialized from D&S and 010 to formalize most of the complex MPEG-7 low-level descriptors and Algorithmic patterns are created to for detection and classification of digital data. COMM covers the "Navigation and Access" part of MPEG-7 that has been intended to use for describing the structure and the content of multimedia documents and is designed for further extension. COMM ontology creates the basis for formalizing and creating annotations of the content of the multimedia document considering its (i) structure and/or (ii) the media as a whole e.g. an Image (media) that realizes ImageData (multimedia content) has been generated by using JPEG compression and that its file size is 273 KB. The COMM ontology covers these two fundamental functionalities by formalizing the decomposition of multimedia content into segments (using Decomposition pattern), or allow content annotation (using content annotation pattern) and media annotation (media annotation pattern). Again to enable annotating with domain specific ontology concepts, content annotation pattern has been further specialized to create semantic annotation design pattern that allows the connection of multimedia descriptions with domain descriptions provided by independent ontologies. COMM is an OWL DL ontology that covers only MPEG-7 MDS and visual descriptors. To achieve full expressiveness COMM needed to be further extended with audio descriptors and OWL 1.1 features.

2.3.4. The MPEG-7 Part4: Audio and associated tools

The Part 4 (Audio) of the MPEG-7 Standard specifies description mechanism as a foundation layer for low-level tools applicable to sound and audio domain in general to create compatibility across audio descriptions. The Audio part has been designed to work with other part of the standard such as Part 2 and Part 5. It assumes knowledge of Part 2: Description Definition Language (DDL) and also has dependencies upon clauses in Part 5: Multimedia Description Schemes, namely many of the fundamental **Description** Schemes that extend the basic type capabilities of the DDL.

MPEG-7 Part 4
Audio

Audio Framework

High Level Tools

Scalable Series
Low level audio descriptors
Silence

Audio Signature
Timbre (type of instrument)
General Sound Recognition and Indexing
(sound model, classification etc)
Spoken Content (Phoneme, word etc)
Melody (Melody, meter, scale, key, beat etc)

Figure 2.12: Overview of MPEG-7- Part 4 Descriptors and Description Schemes

The MPEG-7 Audio part[11] comprises of Descriptors and Descriptor Schemes to define two classes of tools as presented in figure 2.12. The Generic low-level tools known as the (I)Audio Description Framework that contains scalable series, low level descriptors and silence segments designed to provide a basis for construction of (2)Higher Level applications specific tools including general sound recognition & indexing tools, instrumental timbre tools, spoken content tools, audio signature tools and melody description tools.

The audio description framework supports two ways of describing low-level audio features from an audio signal - features extracted from a series of regular intervals or features extracted from arbitrary segments using ***AudioSegment*** to demark regions of similarity and dissimilarity within the sound. Both of these two ways are based on the low-level descriptor types, ***AudioLLDScalarType*** and ***AudioLLDVectorType.*** So, sampled values in a ***ScalableSeries***

11 http://www.tom.comm.waseda.ac.ip/map7/table.html

may be instantiated using one of two ways. Firstly each sample may instantiated from either of the *AudioScalar* or *AudioVector* type, and secondly as a summary descriptor within an *AudioSegment*. *AudioSegment* concept is specified in Part 5 (MDS). An AudioSegment may be decomposed hierarchically to describe a tree of Segments where each segment denotes the temporal interval of audio material, defined by *MediaTime* descriptor that denotes the beginning and end of the segment. The extent of the audio segment may range from arbitrarily short intervals to the entire audio portion of a media document.

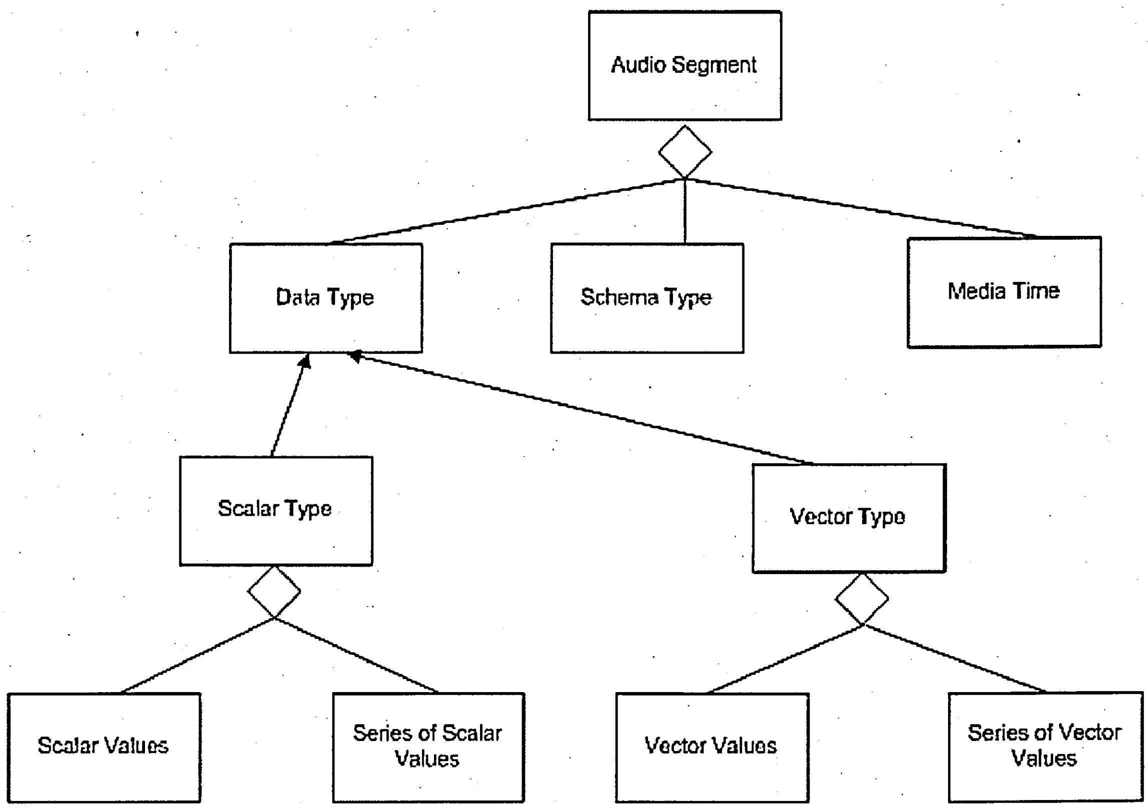


Figure 2.13: MPEG-7 Data Types

Another key concept is in the abstract datatypes: AudioDType and AudioDSType. In order for an audio descriptor or description scheme to be attached to a segment, it must inherit from one of these two types (Scalar type or Vector Type) as described in part 5; a scalar type may be composed of a scalar value or a series of scalar values; similarly in case of vector types - the relationship between these types has been illustrated in figure 2.13.

In MPEG-7 Part 4 there are seventeen low-level audio descriptors that support general audio description. They can be divided into six groups (Lindsay et al. 2002) as shown in the table 5.

Table 5: MPEG-7 Part 4: Low level Descriptors

Data type	Low Level Descriptors category	Descriptors	Abbreviation
Scalar Types	Basic Descriptors (2)	***AudioWaveformType, AudioPowerType***	***AW, AP***
	Basic Spectral Descriptors (4)	***AudioSpecctrumEnvelopeType, AudioSpectrumCentroidType, AudioSpectrumSpreodType, AudioSpectrumFlatnessType***	***ASE, ASC, ASS, ASF***
	Signal Parameter Descriptors (2)	***AudioFundamentalFrequencyType, AudioHarmonicityType***	***AFF, AH***
	Temporal Timbral Descriptors (2)	***LogAttockTime, TemporolCentroid***	***LAT, TC***
	Spectral Timbral Descriptors (5)	***Spectra/Centroid, HarmonicSpectrolCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, FiarmonicSpectralVoriation***	***SC, HSC, HSD, HSS, HSV***
Vector Types	Spectral Basis Descriptors (2)	***AudioSpectrumBasisType, AudioSpectrumProjectionType***	***ASB, ASP***

Two basic descriptors represent instantaneous waveform and power values. Three basic spectral descriptors provide a very compact description of signal's spectral content - reflects approximately the logarithmic response of the human ear represented in dB unit, the perceptual sharpness of the signal and differentiate noise like sounds from tonal sounds respectively.

Among the two signal parameter descriptors - namely the ***AudioFundamentolFrequencyType*** provides an estimated value of the fundamental frequency F0 of the audio signal- represents information regarding musical pitch and the periodic content of the speech signal. Also, it gives indication of melody. The F0 measure may also be utilized with the melody tools described under the high level tools. Two temporal timbral descriptors describe the signal's power function over time within the context of a single well segmented sound. These two descriptors may be used independently or under the high level timbre tool e.g. the *LogAttackTime* is defined as the log(base 10) of the time in seconds from a signal's starting point to the time that it reaches a sustained section (harmonic signal) or a maximum (percussive sound). The four spectral timbral descriptors complement the two Basic descriptors and expand the possibilities by those two. Now, these low level descriptors may be used or combined to indicate musicological property such as timbre or melody of the musical audio.

The **general sound recognition and indexing** tools support for the automatic sound effect identification and indexing. These tools are divided into three different categories. The **spoken content** description scheme is defined for indexing and retrieval of audio stream pertaining to spoken content- which is beyond the scope of this research. In the context of digital music only Timbre and Melody description tools will be utilized to generate the proposed music description vocabulary.

Table 6: Five Levels of Contour Information derived from MelodyContourDS

Contour Value	Change in interval
-2	Descent of a minor third or greater
-1	Descent of a half step or a whole step
0	No Change
1	Ascent of a half step or a whole step
2	Ascent of a minor third or greater

Melody is a successive line of tones or pitches that are characterised by frequency, contour or shape and movement which is structured by its shape & intensity. There are two types of Melody DS tools – one for extremely terse, efficient melody contour representation, and another for more verbose, complete, expressive melody representation.

MPEG-7 **Melody** description tools include *MelodyCountour DS* (for extremely terse melody) and *MelodySequence DS* (for verbose complete melody) for monophonic melodic information. *MelodyCountour DS* uses five step contours (generally termed as shape of the melody) representing the interval difference between adjacent notes ranging between -2 and +2 as shown in the table 6 (Day, 2002). Melody contour or shape of a melody may be categorized as "rising", "falling" or "arch-shaped" melody (Schmidt-Jones, 2010). **Timbre** descriptors are aimed at describing perceptual features of instrument sounds. Perceptual features determine the differences between two sounds with same pitch and loudness to sound differently. The MPEG-7 Timbre descriptors relate to notions such as attack, brightness or richness of sound. In the timbre descriptors, the two widely used classes of musical instrument sounds have been detailed – the harmonic (sustained - coherent) and percussive (non -sustained) sounds. The timbre description tool combines several low level descriptors for harmonic *(HarmonicSpectralCentroid, HarmonicSpectralDeviation, HarmonicSpectralSpread, HarmonicSpectralVariation & LogAttackTime)* and percussive (*SpectralCentroid, TemporalCentroid & LogAttackTime*) sounds.

The **Timbre** descriptors are aimed at describing perceptual features of instrument sounds. Perceptual features determine the differences between two sounds with same pitch and loudness to sound differently. The descriptors relate to notions such as attack, brightness or richness of sound. In the timbre descriptors, the two widely used classes of musical instrument sounds have been detailed – one is the harmonic (sustained - coherent) and the other one is the percussive sounds(non -sustained). The harmonic sounds are described using four Spectral Timbral descriptors and the *LogAttackTime.* On the other hand percussive sounds are described using *SpectralCentroid, TemporalCentroid & LogAttackTime.*

As many of the timbre descriptors rely on a previous estimation of the fundamental frequency and the harmonic peaks of the spectrum or on the temporal signal envelope based on the estimation of spectral parameters, fundamental frequency, harmonic peak and temporal parameters such as *LogAttackTime* and *TemporalCentroid.* Harmonic peaks are determined by the maxima of the amplitude of the Short Time Fourier Transform (STFT) close to the multiples of the fundamental frequency. The frequencies of the harmonic peaks are then estimated by the positions of these maxima while the amplitudes of these maxima determine their amplitudes.

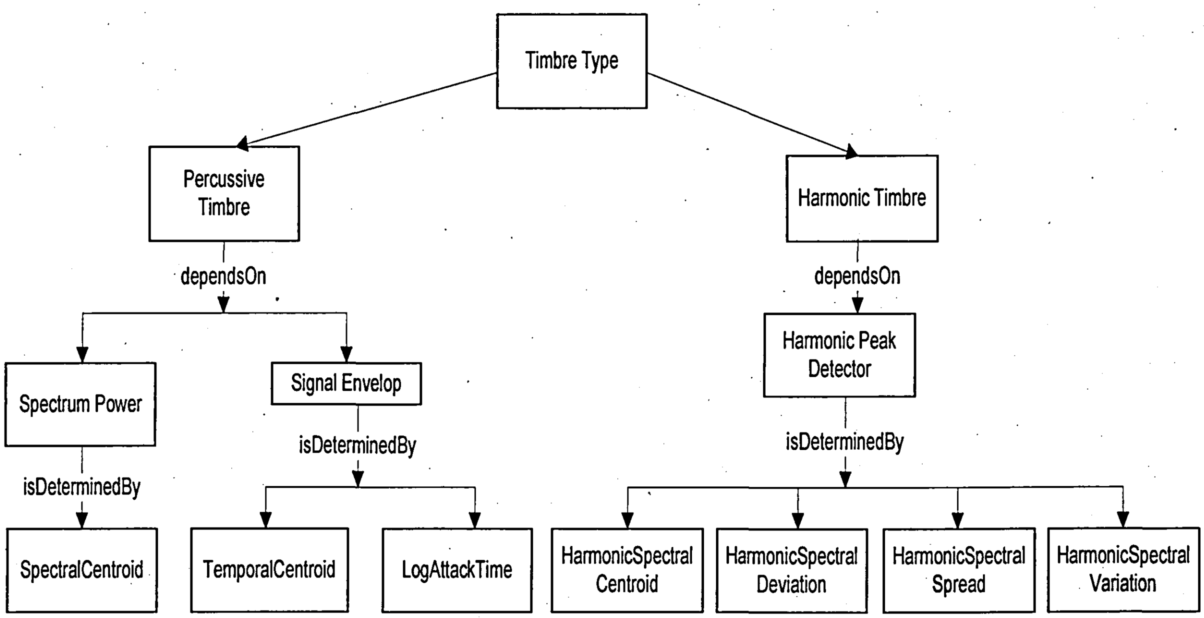


Figure 2.14: Modelling Timbre categories

To summarize on different timbre categories let's have a look at figure 2.14. MPEG-7 Part 4 provides descriptors for two types of timbre – percussive and harmonic. Percussive timbre depends two sets of parameters – spectrum power and signal envelope. Spectrum Power is determined by *SpectralCentroid* which is computed as the power weighted average of the frequency of the bins in the power spectrum. Signal Envelope is determined by

TemporalCentroid (TC) and LogAttackTime (LAT). TC is defined as the time averaged over the energy envelope. LAT denotes the 'attack' of a sound is the first part of a sound, before a real note develops.

Harmonic timbre depends on harmonic peak detector which is determined by four spectral parameters as shown in figure 2.14. *HarmonicSpectralCentroid* (HSC) is derived from experimental results on human perception of timbre similarity using linear scale. *HarmonicSpectralDeviation* is derived using amplitude scale instead of a linear one is derived from experimental results on human perception of timbre similarity. *HarmonicSpectralSpread* is computed as the amplitude weighted standard deviation of the harmonic peaks of the spectrum, normalized by the instantaneous HSC. *HarmonicSpectralVariation* is defined as the normalized correlation between the amplitude of the harmonic peaks of two adjacent frames.

For the purpose of audio descriptioin MPEG-7 provides two methods of aggreagating data (Lindsay et al. 2002). The first one is from the generic multimedia description scheme - an audio stream may be hierarchically decomposed using recursive structure into audio segments where each segment depicts a temporal interval to which all low level descriptors apply. The second one is through the definition of two subtypes of low level descriptors - low level descriptor may be instantiated as a single value for an audio segment or a sampled series using the two types of low level descriptor: Scalar type for scalar values (power or fundamental frequency) and Vector types for spectral content. The sampled values may form a scalable series that enables to progressively down-sample the data contained in the series to store all values of a series as a vector within a raw element. It also stores various summary information such as minimum, maximum, mean and variance of the descriptor. The *AudioLLDScalarType* and *AudioLLDVectorType* allow single value and multi value descriptors respectively. Within these two types, the sampling period namely the hopsize is defined. This thesis followed the second method of aggregation.

Among the Scalartypes the fundamental frequency (AFFT) is a good predictor of musical pitch as well as an important descriptor of an audio signal. This fact will be used to draw meaningful comparisons of musical data labelled with melody descriptor and data labelled with fundamental frequency. Harmonicity (AHT) measure shows distinction between musical sound (harmonic spectrum) and noise (non-harmonic spectrum).

The proposed music annotation ontology will generate appropriate concepts and properties to represent different timbral and melodic class concepts that enable the music producers to derive implicit association with MPEG-7 low level descriptors. This requires extracting MPEG-7 features linked to timbral and melodic descriptors. The next section presents a study of existing available tools that will enable us to generate MPEG-7 features output from music files.

2.3.5. MPEG-7 Feature Extraction Tools

Initially, the search for MPEG-7 feature extraction tool started to find a tool that could generate XML description of music audio files covering all the low level audio descriptors (there are a set of seventeen sound features as discussed in section 2.3.4) automatically from digital musical objects because they represent physical properties of music audio. These descriptors provide a measure of several characteristics of sound represented in the music using XML file format that forms the basis to create advanced MPEG-7 audio description. At present there are several MPEG-7 low level audio feature extraction tools available. The Technical University (TU) of Berlin Audio Analyzer (MPEG-7 Audio Encoder, 2008) implements all 17 audio descriptors defined in the MPEG-7 standard and has created an online web based interface to upload an audio file and generates an XML output file containing MPEG-7 tags and values extracted from the uploaded audio file (supporting formats- .wav and .mp3 and size restrictions < 1MB for .wav and < 300KB for .mp3).

Another tool is MPEG7AudioEncoder[12] by Crysandt (2005) provides Java library that currently extracts the some Low Level Audio Descriptors and description schemes (but not all) from audio files and creates .xml output. This MPEG-7AudioEncoder tool could be a better option to consider but it produces a limited number of low level descriptor which is not sufficient to describe both timbral and melodic attributes of music that are required to formulate the proposed ontology. On the other hand TUBerlin Audio Analyzer is a standalone web based tool to download XML output comprising all the 17 low level descriptors specified under MPEG-7 Audio part (Part 4). So, the TU Audio Analyzer will be used to create XML representation of MPEG-7 descriptors that would be utilized for modelling the proposed ontology. On the otherhand, MPEG7AudioEncoder's (Crysandt, 2005) java plugin will be used to implement the Semantic Annotator Tool that is detailed later in section 3.6. Within the scope of this thesis, the method and algorithms of the audio analyzer by which the feature extraction tools produce the output metadata are not relevant to explain further in this thesis but a further detail can be found in Crysandt (2005).

2.3.6. Summary

The previous section discussed several low level descriptors that may be utilized together in conjunction with the higher level tools to represent different aspects of sounds. The audio fundamental frequency provides an estimation of the melody that can be utilized further with the higher level melody description tools to indicate ascending and descending tones of music. Moreover, Spectral Timbral descriptors and LogAttackTime together may form the basis for categorization of musical sound to harmonic and percussive classes. The existing music annotation vocabularies do not stipulate any way to utilize these facts about musical sound. The basis of this contribution is on the idea of lifting this auto generated concepts of musical sound to the proposed annotation vocabulary by creating a platform for enabling music

[12] Java MPEG-7 Audio Encoder, Java MPEG-7 Audio Encoder 2008. [online] Available at: < http://moeg7audioenc.sourceforge.net/index.html/> [Accessed 8 July 2008]

producers to utilize these characteristics of musical sound. Such attempt will serve at least two purposes. One is the extension of MPEG-7 semantics and other enhancing annotation by music producers and retrieval of music by ordinary listeners – both of these types of users are generally not aware of the underlying knowledge represented by musical sounds.

Current MPEG-7 compliant ontologies started modelling their multimedia ontology starting from the top considering multimedia objects and then the multimedia segments and how these segments are connected with each other and hence serves as upper level concept ontology. These upper level concept ontologies provide generic terms (e.g. AudioSegment) and relations (relation types) that may be further extended for specialized applications but are not sufficient for tasks like semantic annotation by music producers. For example the ABC ontology and MPEG-7 MDS ontology (mentioned in section 2.3.3) provide generic upper level ontology and domain specific upper level concepts respectively to describe MPEG-7 metadata semantics.The AudioSegment concept from ABC ontology can be further extended to MusicSegment to describe features from music domain. MPEG-7 MDS ontology creates upper level concepts to integrate domain specific OWL ontologies with MPEG-7 multimedia description schema.

In contrast to that the music annotation ontology will be modelled from the bottom considering basic signal level information the musical sounds represent specified by MPEG-7 description of music audio. Properties defined in the proposed ontology will create association of MusicSegment class with MPEG-7 audio descriptors characterizing musical properties e.g. timbre and melody.

3- Ontology for Annotating Digital Music

This chapter presents the actual contribution for this thesis; additionally an assessment to the soundness and impact of the contribution is detailed in chapter 4 and 5. Theoretical basis of research methods and evaluation approaches of this contribution are described in section 3.1. Then, it follows the explaination of the levels of music information that has been considered in designing the proposed ontology in Section 3.2. Section 3.3 elaborates the necessary underpinnings for designing the proposed semantic annotation ontology for digital music. Section 3.4 shows the detail design of the main contribution which is the ontology for annotating digital music as well as the simple semantic annotation tool that supports the effectiveness of the proposed ontology and completes the core contribution under this thesis.

3.1. Research Methods

The research process (William, 2006) usually starts with a broad area of interest, but, usually the initial interest is far too broad to study in any single research project. So, it is needed to narrow the question down to one that can reasonably be studied in a research project. This might involve formulating a hypothesis or a conjecture - often referred to as Deduction.

The alternative way for deduction starts by data collection and then the research is conducted first by understanding those data, usually by analyzing it in a variety of ways attempting to address the original broad question of interest by generalizing from the results of this specific study to other related situations - such approach is often termed as Induction.

So, two broad methods of reasoning can be found in research process – known as the ***deductive*** and ***inductive*** reasoning. *Deductive approach* works from the more general to the more specific. Sometimes this is informally called a "top-down" approach. One might begin with thinking up a *theory* about a particular topic of interest and then narrow that down into more specific *hypotheses* that can be tested. Even a further narrowing down may be needed to collect *observations* to address the hypotheses. This ultimately leads to the ability to test the hypotheses with specific data - a *confirmation* (or not) of the original theory.

Inductive approach works the other way, moving from specific observations to broader generalizations and theories. Informally, sometimes this is called a "bottom up" approach. Inductive reasoning begins with specific observations and measures begin to detect patterns and regularities, formulate some tentative hypotheses that can reasonably be explored, and finally end up developing some general conclusions or theories. Inductive reasoning is more open-ended and exploratory, especially at the beginning. Deductive reasoning is narrower in nature and is concerned with testing or confirming hypotheses.

The research method followed in this thesis took an inductive style. It started by looking at the area of semantic search and retrieval of music and considered exploring the strength and weaknesses of music tagging techniques used for/by the music producers because, music tagging is usually encouraged to enable efficient search and retrieval. It was found that music tagging techniques leads to ambiguity for search algorithms as it relies on unstructured tags. The prospect of semantic search algorithms requires to represent music tags with a meaningful (machine process-able) representation. To create meaningful representation for ordinary users' tags different conceptual representation techniques were studied, specifically the use of ontologys for annotating music as well as multimedia representation standard i.e. MPEG-7. We have chosen to design ontology for annotating digital music from MPEG-7 compliant multimedia ontology and Music Ontology. To illustrate the usability of the proposed ontology a simple semantic annotation tool has been designed based on the research on Web2.0 and Semantic Web technologies.

Evaluation of advanced technologies pertains primarily to attributing a measure of the quality of users' satisfaction, and level of the projects' success. Evaluation can be distinguished between traditional, quantitative assessment on one side of the spectrum and authentic assessment, which involves a variety of qualitative approaches on the other side (Dori, 2007).

There are many different types of *evaluations* depending on the object being evaluated and the purpose of the evaluation. Perhaps the most important basic distinction in evaluation types is that between ***formative*** and ***summative*** evaluation (William, 2006). Formative evaluations strengthen or improve the object being evaluated - they help form it by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object - they summarize it by describing what happens subsequent to delivery of the program or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

Simon (1969) established the foundation for **Design Research** by dividing the body of knowledge into two distinct categories (natural science vs. artificial science) to bring design activity into research. The body of natural science is concerned with objects or phenomenon of nature or society and explains how they behave and interact with each other. On the other hand the body of knowledge concerning artificial science that describes man-made (artificial) objects and phenomena designed to satisfy certain number of goals. Design research provides a lens or set of analytical techniques and perspectives for performing research in Information Science & Technology or Software Engineering or Computer Science involving performance analysis of designed artefacts to understand, explain and improve on the behaviour of aspects of Information Systems. Such artefacts include information retrieval applications, human/computer interfaces, game design, and algorithms but certainly not limited to algorithms only. In a widely cited work by March and Smith (1995) contrasted design research with natural science research and proposed four general outputs for design research termed as constructs, models, methods and instantiations.

Constructs are the conceptual vocabulary of a problem/solution domain and arise during the conceptualization of the problem and get refined throughout the design cycle. A **model** is described as "a set of propositions or statements expressing relationships among constructs" in March and Smith (1995). Models differ from natural science theories primarily in intent. Traditionally natural science focuses on truth and tries to understand the reality. In contrast, design research focuses more on things created to serve a particular purpose and assessing situated utility. Thus a model is presented in terms of what it does and a theory described in terms of construct relationships. A **method** is a set of steps (an algorithm or guideline) used to perform a task. An **instantiation** actually "operationalizes constructs, models and methods" (March and Smith, 1995). It is the realization of the artefact in an environment.

This thesis planned to conduct design research to design and build the proposed semantic annotation ontology (which is the 'construct') for digital music with the detail design to enable music producers for annotating digital music as a model of semantic annotation of digital music by music publishers. It has designed and developed a simple semantic annotation tool showing detail steps to perform annotation as a method to support music producers to perform the task of annotation of digital music. Also, the annotation tool has been designed to act as a foundation to support the effectiveness of the proposed ontology and it shows the instantiation of the proposed music annotation ontology construct.

The two basic activities of design science are - build and evaluate. Building activities stipulate the process of constructing the artefact and produces outputs like constructs, models, methods and instantiations as solution towards solving a specific problem. Then evaluation activities perform the process to assess the criteria of value or utility of those solution e.g. user acceptance field trials etc.

Research in design science is technology oriented and its outputs are evaluated in relation to intended use of the artefact to perform certain tasks. So, to evaluate this contribution based on advanced technologies a summative evaluation of the proposed outcomes has been carried out and a critical evaluation has been presented detailing its impact in the related research fields mentioned in figure 2.1.

3.2. Representation levels of Music Information

Music unfolds in time and hence it has got tempo or duration. Complexities of human perception of music are related to the temporal aspects such as beat, rhythm and tempo. Again, music is conventionally described using terms like melody (Schmidt-Jones, 2010), harmony, rhythm, dynamics, timbre.

Melody is a successive line of tones or pitches that are characterised by range, shape, movement or motion. Timbre is the characteristic sound of an instrument or voice that enables to differentiate the sound from other instruments when they are playing the same note. Sound produced by conventional musical instrument or voices are usually not sinusoidal rather they are mixture of sine waves. This non-sinusoidal character is related to the ***timbre.***

Only melody and timbre have been used to formulate concepts of the proposed vocabulary. Because, the tool used to generate MPEG-7 XML output only extracts melodic and timbral descriptors from music files. It is possible to incorporate other dimensions such as tempo but it requires using tempo counters that would auto generate XML output conforming to MPEG-7 Part 4 syntax. I have left it for future improvement of the proposed vocabulary. A detail of all other musical dimensions have been added in Appendix A for further reference.

With regard to music search and annotation the goals set out in section 1.2 synthesized the objectives of this thesis to address one research question which was "How to annotate music against MPEG-7 description to deliver meaningful search results?" As a solution towards providing meaningful search results to music consumers, the literature study of existing music tagging techniques showed that those techniques were aimed towards generating music metadata by music consumers so that search algorithms manipulating those could yield efficient search results and retrieval. But, the limitations of those tagging schemes did arise from unstructured representation of generated metadata (section 2.1). Then closer look was focussed towards how music could be semantically annotated so that generated annotations would produce structured meaningful metadata by providing machine process-ability. The methods and technology available were studied for structured, meaningful, conceptual representation of metadata to describe multimedia (specially, semantic web technologies and tools) as well as standard metadata schemes to describe multimedia (i.e. MPEG-7) in section 2.2 and 2.3 respectively. Also, annotation tools available (i.e. tools developed using Web2.0 technologies) were studied so that one of those could be utilized or adapted to annotate music content with structured metadata vocabulary designed and represented using semantic web standards like OWL. Then the proposed ontology was planned to be developed as a structured metadata to annotate music by music producers. Before, detailing the essential building blocks and modelling principles of the proposed ontology, different levels of information conveyed by a piece of digital music will be presented next.

There are four types of music representation found in the existing literature and these types are organized in levels associated to the various information quantity they convey (Vinet, 2004). Figure 3.1 depicts four level of musical information associated with digital music. The lowest level physical layer can be characterized by the acoustic pressure as a function of space and time i.e. the acoustic pressure signals at the level of both eardrums, which characterize information inputted into the auditory system as played back from digital music files.

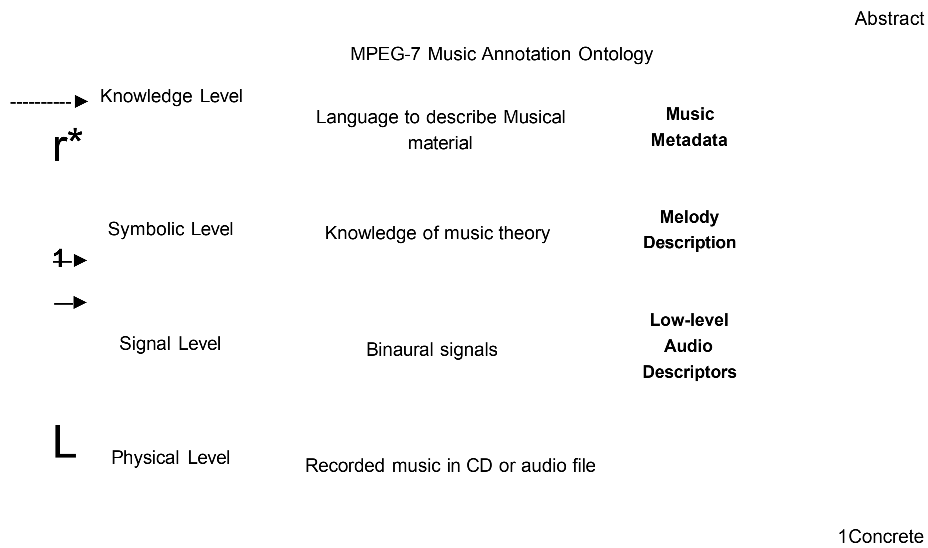


Figure 3.1: Representation levels of Music Information-adapted from Vinet (2004)

The signal level representation specifies a content-unaware representation to transmit any sound (both musical/non-musical/audible and even non-audible signals). These signal level representation of music can be automatically extracted from music file using MPEG-7 feature extractor tools (the detail was presented before in section 2.3.5).

The symbolic representation describes ***content-aware*** events with regard to the formal concepts of music theory and accounts for discrete events, both in time and in possible event states (e.g. melody shape & motion, musical timbre) with reference to listeners' knowledge of music theory. For instance, musical timbre may be described as multidimensional perceptual attributes that comprise relevant information that conveys the identity of the sound in a musical context. Typically spectral, temporal attributes and acoustical parameter of sound are used to determine timbre such as spectral envelope, spectral centroid, temporal envelope and spectral flux respectively (Donnadieu, 2007) and creates a relation to verbal attributes used to describe timbre perception of complex sounds e.g. smooth-rough or light-dark by correlating verbal attributes with one or more perceptual dimensions - "dry" correlated with the log of the attack time dimension, "round" correlated with the spectral centroid, "brilliant/bright" correlated with spectral centroid while "metallic" was correlated with three perceptual dimensions.

Music consumers with little or no understanding of musicology/music theory might be able to use simple phrases at a maximum e.g. rising, falling etc. to describe the musical item and may not be able to properly classify the part or whole music according to the melody (Schmidt-Jones, 2010). Similarly, timbral (tonal aspect) features of music may be named using everyday words like having gentle, metallic, hard sound characteristics by the listeners (Sarkar et al., 2007). But, these musicological features are understood in the domain of music publishers/producers who are able to classify correctly a musical item using melodic and timbral dimensions. Besides, the MPEG-7 Audio standard (detailed in section 2.3.4) provides description schemes based on signal-level audio descriptors that represent musical timbre and melody and these MPEG-7 low-level descriptors automatically generated from music files (using freeware tools as mentioned in 2.3.5) may be mapped to symbolic level concepts (timbre and melody).

The knowledge level representation is associated with appropriate language structures for describing musical phenomena. This thesis was aimed to create a meaningful representation of music metadata using conceptual representation languages as detailed in section 2.2.2. Using highly interoperable and semantically rich languages musical dimensions may be modelled to create knowledge level representation. But this would require establishing mapping among the different levels first.

Mapping from signal to symbolic level is associated to a *double digitization* concerning both the *time axis* and the *value ranges* taken by analyzed variables e.g. the fundamental frequency is extracted from audio signals as a low bandwidth, slow variation signal, then it is again quantized over time into note events, and the frequency values are mapped into a discrete semitone pitch scale. Information conveyed by symbolic level pertains to musicological concepts like melody, timbre etc. that are related to different dimensions of music (the detail is in Appendix A).

Knowledge level representation may be achieved from both signal and symbolic levels (as shown using arrows in figure 3.1). Signal level information that specifies a set of audio descriptors may be mapped to represent objective representations of music items. Symbolic level Melody and Timbre Description schemas may be conceptually modelled to provide subjective representation of musical items. So, knowledge level representation of music material provides both objective descriptions (mapped from MPEG-7 low level audio features related to timbral and melodic description schema) and subjective descriptions (features modelled from musical dimensions e.g. timbre and melody). As a result, knowledge level representation can contain objective and subjective characteristics that can be created from combinations of signal and symbolic level information respectively.

The limitation of such layered representation is its inadequacy to reflect the structural complexity of the intermediary levels and to model the way higher-level information is structured. To overcome the problem of conversion between various representation levels and to adapt the different level of information to support technical applications such as search and retrieval of music it was proposed earlier in this thesis to create a semantic vocabulary to support the process of annotation of music files by music producers. This vocabulary (that is going to be detailed in the next two sections) has been designed to provide a set of formal concepts and properties as an abstraction model for signal and symbolic level information.

So, having characterized and scoped the knowledge level representation of musical items containing signal level MPEG-7 audio descriptor metadata (objective characteristics) and symbolic level musical dimensions (subjective characteristics) related to timbral and melody description metadata to create a structured metadata vocabulary for music annotation the essential underpinnings is now being outlined in section 3.3.

3.3. Essential underpinnings of Semantic Annotation Ontology for Digital Music

3.3.1. Introduction

The value of any information resource in computational terms depends on the ability to locate, retrieve and manage those resources (Sonera White Paper, 2003); and for digital music available on the web such characteristic becomes crucially important for music consumers. To enable easy access to any digital media content the MPEG-7 developed standardized ways to describe multimedia and it aims to create interoperable description for multimedia resources but such multimedia description needs to be enriched with meaning that will provide useful response to search query. The big problem with MPEG-7 is that it does not provide guidance on how to formulate a query to satisfy search request (Staab, 2006); for example digital music's audio content.

The MPEG-7 audio description techniques consist of a set of structured XML elements forming types such as ***basicTimePoint,*** basicDuration, ***audio fundamental frequency*** etc. that can be used to identify and represent acoustic features of digital music content. But, these XML elements are adequate only to form a syntactic representation for underlying audio of the musical content. It does not stipulate any tool or hard and fast rule regarding how these XML based descriptions may be further enriched with semantics that can be used to match search queries for musical objects. These syntactic descriptions of musical contents (using XML) require further lifting towards semantically enriched machine interpretable descriptions (Ossenbruggen et al., 2004). But further lifting of MPEG-7 description requires us to draw a mapping and association with semantic concepts; using standard semantic representation paradigm such as Web Ontology Language (OWL) that will preserve the interoperability aim of MPEG-7 initiative.

The main focus of this research effort for the purpose of developing annotation ontology for digital music is towards the utilization of MPEG-7 syntactic metadata created from audio signal encoded in the music files. To extract audio signal's syntactic descriptors from music files available tools were examined; that were designed to extract those descriptors according to the syntax specified by ISO/IEC 15938 Part 1 (Martinez et al. 2002). Then to model the annotation ontology according to the set of requirements (that has been mentioned later in section 4.2.1) existing MPEG-7 ontologies were analyzed i.e. ABC ontology (Hunter and Little, 2007) to determine the possibility of reusing or adapt them to support the proposed annotation ontology. Also, several existing music metadata standards as well as the music ontology were studied i.e. Raimond's Music Ontology (Raimond et al., 2007) that may serve as a foundation of the Music Annotation Ontology. Design of the proposed ontology also required to study the state-of-the-art research on how musicians and music lovers perceive different aspect of musical pieces as well as how audio signal level data may be projected to conceptual aspects of music producers understanding of music so that the resulting ontology could adequately provide the knowledge representation foundation.

The subsequent sections will elaborate on the type of music search queries, manipulation of MPEG-7 feature extractor's output, existing MPEG-7 ontology etc. because, the design of the proposed ontology involves two stages development. First, mapping of the MPEG-7 low level audio features has been done as they have to be linked to timbral and melodic characteristics of music. This prompts us to examine input queries related to timbral and melodic features that are not currently supported by any tagging techniques and so by search engines. Then, available options to manipulate XML output of MPEG-7 features need to be explored. Once those MPEG-7 metadata are available for use from XML data source those have to be linked through definition of appropriate property to map to the proposed semantic annotation ontology.

As mentioned in section 2.1.3, the existing music tagging techniques suffers from low precision & recall and cold start problem, because the pay attention only to general purpose tagging to support queries like (Baumann et al. 2002):

a. Find Music by Madonna (find by artist name only)
b. Find Music by Madonna made between 1985 and 1995 (find by artist name and physical time context)

Apart from these, the following queries seem to be trying to find out different musical pieces by same artist but narrow the search using both semantic and acoustic properties e.g.

i. Find Music by Madonna made between 1985 and 1995 but slow (find by artist name, physical time context as well as acoustic tempo)
ii. Find Music by Madonna made between 1985 and 1995 but percussive (find by artist name and physical time context and acoustic timbre)
iii. Find Music by Madonna made after 2006 but faster (find by artist name and physical time and acoustic timbre)
iv. Find Music by Madonna made that has rising melody (find by artist name and physical time context and melodic feature)

Present music metadata schemes (as mentioned in section 2.1.4) do not provide sufficient tags for content based query as shown above (i-iv). At present, no search engine is able to return satisfactory search results for this type of queries. Even users (both music producers and consumers) in general are not equipped with any structured metadata schemes to tag/search music with content based features like timbre or melody (mentioned in section 3.2). So, in the design of the proposed ontology provision for music producers to annotate music with timbral and melodic features has been incorporated as represented in queries ii-iv above.

3.3.2. Using MPEG-7 XML Metadata

This section will detail the role of MPEG-7 audio metadata in the design of the proposed vocabulary. In order to actually utilize/extract XML output of MPEG-7 feature extraction tools, this work initially considered different ways: firstly the Rhizomik XSD20WL utility[13] (Garcia and Celma, 2005) that uses XPath processing to generate OWL ontology from semantic relations present in the XML schema. Use of such conversion utility is usually very straightforward and effective but then it would require aligning the converted OWL ontology with the proposed annotation ontology; which requires further research on how to align ontologies. Second option was the use of a customized tool to extract the chosen MPEG-7 descriptor (s) from the XML output.

This thesis has chosen the second option of creating a XML data source processor analyze MPEG-7 XML output because the first option requires us to pursue ontology alignment which is another broad area of research apart from developing the proposed ontology. Using MPEG-7 XML description as data source needs to examine the logical structure of XML data. The logical structure of XML data is modelled as Document Object Model (DOM) (Luo et al., 2004) to manage, access and manipulate the data presented in the XML documents. DOM is a language independent application program interface (API) that can be used in any environment. The model is a tree like structure (actually may contain more than one tree). So, DOM parses an XML document and returns an instance of the W3C.org's Document object

[13] XSD20WL Utility, available from http://rhizomik.net/html/redefer/

(Hégaret, 2008) that must then be walked in to process different elements contained by the XML source.

The remaining task is to choose API for XML processing to manipulate the MPEG-7 feature extractor output. There are several (Hunter, 2000) open source APIs are available for parsing XML data source. SAX (Simple API for XML) was the first one that came for XML processing providing basic sequential support to access the document. DOM is a Java binding for DOM that provides a tree-based representation of the XML documents - allowing random access and modification of the underlying XML data. It replaced the event-based call-back methodology of SAX by an object-oriented in-memory representation of the XML documents. But it is slower than SAX. Both SAX and DOM requires XML and Java knowledge to work with. JAXP (JAVA API for SML Processing) were evolved to reduce the dependency of Java developers from massive XML parser output processing to obtain DOM and SAX compliant parsers with few factory classes to make sure that an interchange between the parsers required minimal code changes. JDOM is an open source API, designed specifically for Java programmers, that represents an XML tree as Elements and Attributes. JDOM provides JAVA API and Java constructs to process XML documents and eliminated the need for Java programmes to use non-Java constructs like Attributes (in SAX) and NamedNodeMap (in DOM). Such approach helps to achieve faster processing speed similar to SAX. Choice of the specific parser technology depends on the requirements of the application. If the entire document needs to be represented, then most likely option would be DOM, JDOM or JAXP's builder implementation. If only parts of the XML document and/or parsing the document happens only once, it might be better served using SAX or JAXP's SAX implementation. Again depending on the programmers comfort and processing speed requirements SAX (XML programmers) or JDOM (Java Programmers) may be chosen.

This thesis have chosen to use JDOM to parse MPEG-7 encoder output because to evaluate the proposed ontology it has been planned to use a customized annotator application which has been developed using Java based technology (JDK 6). As JDOM is specifically designed for JAVA programmers it appears a flexible option to integrate the semantic annotator application to interface with the XML data source processor.

Let's focus our attention to mapping of MPEG-7 XML representation to the proposed ontology. I used the idea of data type mapping rule (Ghawi and Cullot, 2009) that considers datatype properties to emerge from attributes and from simple types. The simple types are mapped to a datatype property having as domain the OWL class corresponding to the surrounding complex type, and having as range its XSD datatype. For example, in the proposed ontology, the MPEG-7 Timbre descriptors (Spectral centroid, Harmonic Spectral Spread etc.) have been mapped to be as sub-property of mpeg7DataType property that used xsd:string as range and MusicSegment as domain. Attributes are treated as simple elements and they have been mapped to datatype properties. If a complex type is mixed, then the elements that have this type contain text as well as sub-elements and/or attributes.

3.3.3. Limitations / Current State of MPEG-7 Ontology

The ABC ontology (Hunter, 2003) actually forms the foundation of MPEG-7 data model to further extend any MPEG-7 ontology. It creates the basic concepts of how the class and property hierarchies and semantic definitions should be derived from MPEG-7 Descriptors (Lindsay et al., 2002) and Description Schemes (MPEG-7 Part5, 2003). The approach is to describe a core subset of multimedia content entities (StillRegion, AudioSegment, VideoSegment, AVSegments etc.) as subclass of the top level MultimediaContent class which was derived from the Resource class. Visual features and properties such as Colour, Texture, Motion and Shape are only applicable to visual entities only; but these are not relevant not for digital items containing solely musical audio. ABC ontology provides owl class hierarchy and property definition of these visual features but does not elaborate on class hierarchy and property definition of audio features. It creates insight on how a similar approach to audio features such as silence, Timbre, Speech and Melody may be applied to MPEG-7 entities that contain audio. The proposed ontology has created the MusicalSegment class by extending the AudioSegment class of ABC ontology.

This thesis has also explored the multimedia ontology proposed by (Hollink, 2006) that extends visual feature Colour (Colour class from ABC ontology) to annotate visual resources for efficient retrieval. Hollink's visual ontology creates subclass of *Colour* to represent dark and light colour and provides associated property definitions. In a similar fashion the proposed ontology has created the MusicalExpression class and its subclasses - TimbralExpression/MelodicExpressions with appropriate property definitions that connect MPEG-7 timbral and melodic descriptors with semantic concepts.

The OWL –DL representation of full MPEG-7 MDS as proposed by Tsinaraki et al. (2004) is aimed to serve as an upper ontology for the ontologies developed to capture MPEG-7 semantics. Their approach was focussed towards developing an upper level ontology that forms the OWL based representation of MPEG-7 data types as OWL: DataTypeProperty, creating OWL class for each MPEG-7 complex types and for every simple attribute of the complex type a datatype property is defined using the owl: DatatypeProperty construct and complex attributes are represented as OWL object properties, which relate class instances. No doubt that, development of such an ontology will ease the effort of creating new MPEG-7 ontologies for different application domain but it does require the development mapping of domain ontology concepts to its upper ontology classes to create annotation for further retrieval of search results, the upper ontology is based on a one-to-one mapping for further extension.

The common issue that all prominent MPEG-7 ontologies do not specify explicitly is the audio features' representation specifically. This thesis has created the extension of the audio feature ontology for the purpose of annotation of music resources that is actually a semantic annotation ontology for digital music based on a set of requirements (detailed later in section 4.2.1) guided by the Common Multimedia Ontology Framework requirements (Saathoff et al, 2006; Eleftherohorinou etal. 2006).

Now, let's turn to the metadata used in the domain of music production as well as music search, retrieval and recommendation in order to identify potential metadata standards used in those domains with a view to showing that how the proposed ontology has been made compatible to those standards. In this context, a study on strengths and limitations of existing music metadata schema and ontology will be presented in the next section.

3.3.4. Existing Music Metadata schema and Ontology

The study of survey of existing metadata standards by Corthaut et al. (2008) that presented a comparison of expressiveness and richness of different music metadata schemas and their relation with application domains shows useful insights about which schemas are relevant for certain application domains. The life of musical piece starting from inception has created different application domain such as music production/playback, music library/management, and commercial transaction of musical items, creation and manipulation of musical notations, music, search, retrieval and recommendation. At present there is no single metadata standard available for music covering all the possible requirements for aforementioned applications of music.

The available music metadata schemas and vocabularies are either focused on interoperability or designed as a standard from the start. The ID3 (Oneill, 2009) tag contains the artist name, song title and genre to be embedded within the audio file and has got a wide spread use in music players and devices e.g. iTunes, iPod and Winamp. These are general purpose syntactic tags and are designed to organize music items for commercial exploitation of music rather than supporting music consumers' search query and do not enable music producers to annotate music with content based semantic tags.

The Kanzaki Music Vocabulary (Kanzaki, 2003) or Kanzaki ontology presents concepts to describe classical music and performances in order to distinguish musical works from performance events or works from performer in the ensemble defining classes for musical works, events, instruments, performers and relationships between them. Based on this vocabulary, the Music Ontology (Raimond et al., 2007) deals with music-related information on the Semantic Web, including editorial, cultural and acoustic Information and has been developed as a ground for more domain-specific knowledge representation. The goal is to enable a "blogger" to put online a recording of a concert he attended the day before, or a musicologist to express complex tonality of musical piece and has been represented in several levels of expressiveness. The music ontology[14] covers mainly the applications related to editorial musical info (albums and tracks), event related concept concerning work flow involving the composition of a musical work, an arrangement of this work, a performance of this arrangement and a recording of a performance. Though the Music ontology creates semantic metadata to incorporate several levels of musical information including editorial information, event related concepts (e.g. arrangement and recording performance) and musicological information (e.g. which key was played at a certain time by person playing the instrument). But, the set of semantic concepts provided by Music ontology do not cover all layers (from physical to knowledge level as in figure 3.1) of information as contained by a musical object that can be utilized for annotation as well as further retrieval.

Yet, Music ontology imports the time line ontology to cover both acoustic signal level aspects and universal physical time may be further extended to accommodate perceptual features associated with the timeline. To accommodate both Physical time and MPEG-7 media time in relation to time instant and duration aspects of audio files the proposed ontology has

14 http://musicontologv.eom/#sec-evolution

borrowed time line concepts by adding two disjoint subclasses (MediaTime and PhysicalTime) to it and appropriate properties have also been created. A further detail will be presented in section 3.4.5.

Earlier in section 3.3.2 the role of MPEG-7 metadata was presented in mapping low level descriptors in the proposed ontology as well as how to manipulate those MPEG-7 features. Now, the methodology to describe the proposed ontology will be presented in the next section.

3.3.5. Methodology to describe the ontology

This section presents the detail of how MPEG-7 metadata is mapped to the proposed ontology. This thesis adopts a two step method to map MPEG-7 metadata to the proposed ontology - the mapping process first performs the linking of MPEG-7 descriptors as data type properties and secondly it turns to modelling of musical concepts in relation to these data type properties. Finally, it focuses on the semantic representation language to encode the proposed semantic metadata model.

Mapping MPEG-7 descriptor as Datatype Properties

Fundamental to all varieties of music is the sound as a physical phenomenon that can be considered both objectively and subjectively (Open University, 2007). Objective phenomena of sound can be measured automatically (e.g. MPEG-7 low level description) and described mostly using scientific vocabulary from physics where as subjective phenomena (e.g. related to melody and timbre) are experienced as a perception using metaphorical description mostly from musical terminology as previously described. As explained in section 3.2 that in the context of contribution of this thesis it considers MPEG-7 low level audio description metadata (related to timbral and melody description schema only) as the objective information; whereas melodic characteristics (like motion, shape) and timbral characteristics (like sharpness, brightness) are considered as subjective phenomena.

It is quite difficult for ordinary people to describe music effectively. Music professionals use different terms like timbre, texture, tempo etc. to represent complex musical dimensions (Appendix A). Linking music with human feelings, mood and understanding has created dispute among musicians and critics (Storr, 1997). In reality, many ordinary listeners appreciate musical forms and structures without being able to describe it using technical and musicological language. If they would not be able to appreciate it then music would not continue to be important to them. Appreciating musical forms and structure is not a technical matter which only trained musician is equipped to undertake and also it should not be dependent on the listeners' scientific understanding of acoustics.

The MPEG-7 low level audio tool specifies audio features describing spectral content and temporal info such as amplitude, envelope etc. A recent survey conducted by MIT Media lab (Sarkar et al., 2007) indicates that musicians and music lovers assign similar words to certain types of sounds, that leads to a mapping between words and timbre and these words are strongly related with audio features both in the frequency and time domain. The melodic and rhythmic information may be transcribed into melody contour calculated from MPEG-7 descriptor *AudioFundamentalFrequency* using five contour values as specified in MPEG-7 Audio standard (Batke et al. 2004).

Based on the fact that specific audio descriptors may be combined (e.g. low level features and signal structure, models and semantics) to map them to application specific semantics, we start modelling the Music annotation ontology to combine timbral descriptors that is easy to map the characteristic sound of the musical piece to ontological concepts namely harmonic and percussive timbre. In order to capture the semantic info originating from different segments of a musical piece the proposed annotation ontology considers segmentation of audio and extends the timeline concept of Music Ontology by defining subclasses like MediaTime and PhysicalTime that handles the decomposition of audio media in time and universal time respectively. Figure 3.2 shows the core class hierarchy that has been created marked clearly using orange ellipses. Class concepts marked with blue and light green ellipses were imported from the ABC ontology and Music Ontology respectively.

Modelling Musical concepts from MPEG-7

The proposed ontology considers segmentation of musical audio by extending the AudioSegment class (from the upper level ABC ontology) and creates MusicSegment subclass.

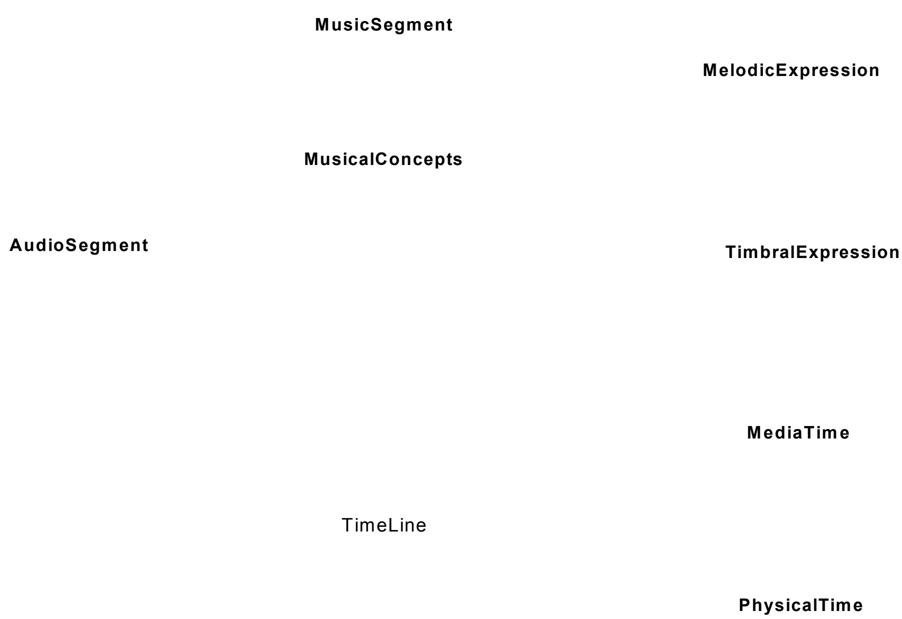


Figure 3.2: Core Classes of Music Annotation Ontology

It also has borrowed the idea of the Interval and Instant class from the Music Ontology[15] but has modelled those as data type properties namely Duration and TimePoint to establish relationship among the MusicSegment and the TimeLine concepts as shown in figure 3.3. The rationale behind modelling interval and instant as distinct datatypes is to incorporate

15 http://musicontology.eom/#sec-evolution

the MPEG-7 ***basicTimeDurotiion*** and ***basicTimePoint*** to describe Music segments. Besides, the TimePoint property associates individual instances of TimeLine class with the MusicSegment.

MelodicDescriptor and TimbralDescriptor are modelled as datatype properties to contain MPEG-7 descriptor from MPEG-7 Audio tools such as ***AudioFundamentalFrequencyType*** (AFFT) and ***SpectralCentroid*** (SC) respectively.

Mpeg7DataType
Descriptor

Duration
Domain: MusicSegment
Range: Float

MelodicDescriptor
Domain: MusicSegment
Range: int
Allowed values for contour: -2=<x<=2

TimbralDescriptor
Domain: MusicSegment
Range: String

TimePoint
Domain: MusicSegment, TimeLine
Range: DateTime

Figure 3.3: Data Type Properties of Music Annotation Ontology

Choice of Semantic representation language

The main reason behind the limitations of current music search and tagging techniques are that these techniques as mentioned previously operate on a syntactic level rather than a semantic one (Yang, 2007). To overcome this limitation the W3C Semantic Web Activity (Herman, nd) establishes standard ways to represent machine process-able web contents to allow machines to follow links and facilitate the integration of data from many different sources. When the relationships among data are fully accessible to machines, machines will be able to help us browse those relationships and interpret the data as well as assess the appropriateness of the data for the intended purposes of users (Miller and Swick, 2003). Thus W3C plays a leadership role in both the design of specifications and the open, collaborative development of technologies focused on representing relationships and meaning and the automation, integration and reuse of data. One of the standard ways of representing data semantics through explicit relationships among different concepts is using 'ontology' that may be used by automated tools to empower more accurate web search. To that aim the Web Ontology Working Group's standard efforts (Connoly et al., 2004) has designed another language to build upon the RDF language namely Web Ontology Language – best known as OWL (Herman, 2007), for defining structured, web-based ontologies. The Semantic Web really started from metadata that provides a syntactic description of data represented in XML. With the development of a resource description framework or RDF that models things and relationships it provides semantic (meaningful description of) data; which is much more reusable and enable machine processabality (Hardin, 2005). With the advent of OWL on top of RDF has brought richer representational capability to model knowledge to further inference on the meaning of data.

As mentioned in section 2.2.2, OWL 1.0 has got three variants (namely OWL-Lite, OWL-DL and OWL-full) with differing expressive capability as well as inference ability. Generally, the richer express-ability leads to the more inefficient reasoning support as well as computability (referring to figure 3.4). So, trade-off is required between express-ability and reasoning support when we choose an ontology representation language.

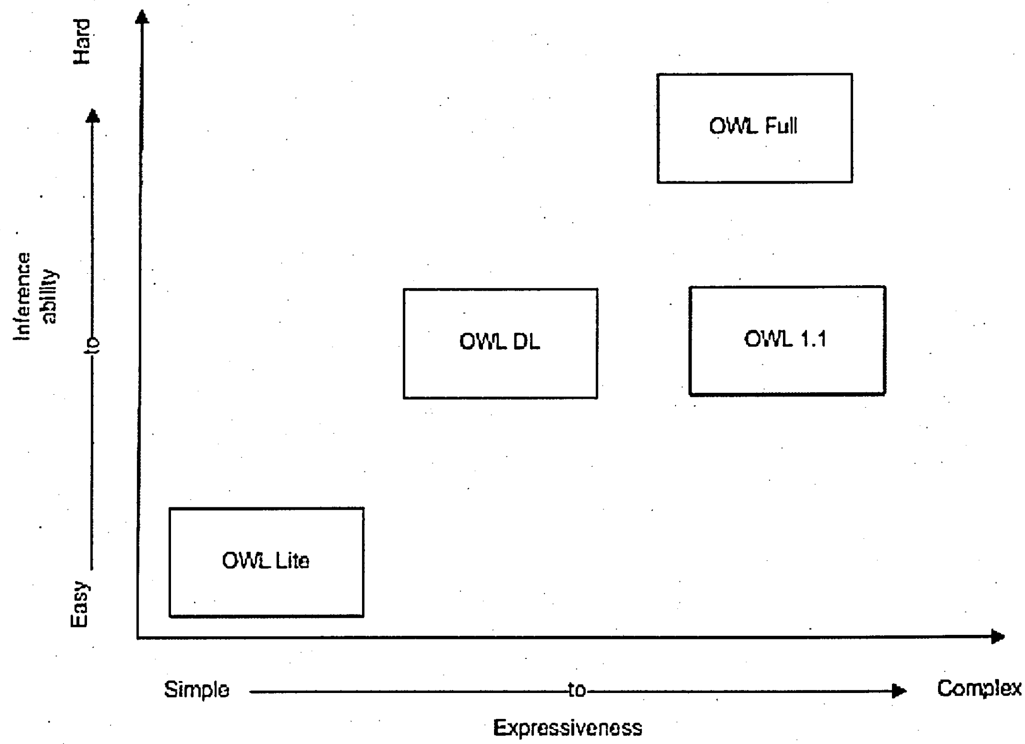


Figure 3.4: Different Species of Web Ontology Language (OWL)

OWL-DL allows expressing 'disjointness' on concepts (classes) but not on roles (properties). But role disjointness can generate new subsumptions or inconsistencies in the presence of role hierarchies and number restrictions. e.g., the roles sisterOf and motherOf should be declared as being disjointed as shown in figure 3.5. So, OWL 1.1 improves OWL-DL by creating an extension of logical basis of OWL-DL ontology language with disjoint, reflexive/irreflexive roles and negated role assertions added to it.

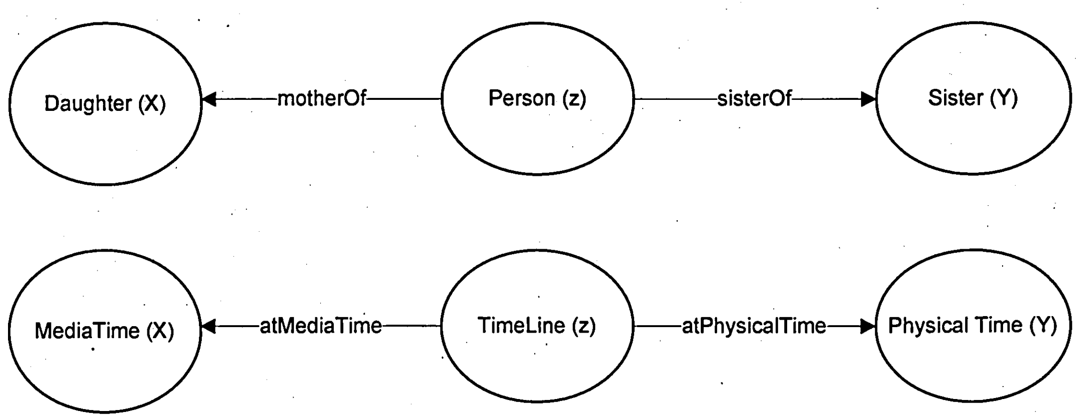


Figure 3.5: Role disjoint Properties of Music Annotation Ontology

The TimeLine concept in the proposed ontology to represent MPEG-7 audio temporal feature may describe two types of time related data, one to denote media data and the other to denote the time with respect to the universal physical time. The roles of atMedia Time and atPhysicalTime needs to be declared as disjoint to maintain consistency (refer to figure 3.5). As we need an ontology that has got richer express-ability as well as inference capability using OWL 1.1 would be ideal for the purpose of modelling the music annotation ontology. The open source tool Protege (version 3.4) has been used to develop this ontology to incorporate association rules (further detail will be presented in section 3.4.5). At the time of developing of this ontology only Protege version 4.0 could support OWL 1.1 features but version 4.0 did not provide rules editor which was available to Protege version 3.4 only. So, the proposed contribution in this thesis encodes the music annotation ontology in OWLI.O format.

3.3.6. Summary

The section 3 has presented the core concepts and properties needed to map MPEG-7 descriptors in order to lift it to the proposed music annotation ontology. It has extended concepts from ABC and Music Ontology and added the MPEG-7 descriptors as datatype properties. To adopt the coding standard for the proposed ontology the ideal candidate would be OWL 1.1 to take advantage of disjoint roles feature but due to the plugin limitation of the Protege tool's version the proposed music annotation ontology has been encoded using OWL 1.0 standard.

To summarize, this section establishes the mapping of MPEG-7 descriptor by introducing the core class concepts as well as datatype properties modelled to lift the MPEG-7 descriptors' objective data as primary and essential components of the proposed ontology.

3.4. Development of the Proposed Ontology

The previous section detailed the mapping of MPEG-7 descriptors in the proposed ontology and from now on in this section the detail development of the proposed ontology will be specified and I will call it as the ***mpeg-7Music*** in short. For clarity, the ***mpeg-7Music*** ontology is the proposed ontology for annotation of digital music in this thesis.

3.4.1. Modelling the Music Annotation Ontology

The construction of the proposed ontology named ***mpeg-7Music*** w 'Wbe specified based on the primary ontology model of the proposed vocabulary of metadata consisting of class concepts and properties as well as their corresponding chosen elements from MPEG-7 descriptors mentioned in the previous section (3.3). The final ontology model will evolve from several perspectives. The primary model of the semantic annotation ontology were adapted from existing MPEG-7 based multimedia ontology and Music Ontology to represent the domain of music files using the audio features and the semantic concepts & properties. Basically, the primary model is extended and reorganized the Music Ontology by adding features to incorporate MPEG-7 High level concepts such as Timbre, Melody and provision for incorporating data types such as time instant (***basicTimePoint***) and time interval ***[basicTimeDuration)*** and it starts from extending ABC ontology's ***AudioSegment*** as root concept. Now, in this section I will describe two fundamental aspect of the proposed ontology: firstly, how we can incorporate music consumers' perspective in the annotation process of music items by music producers and secondly, how the low level MPEG-7 feature may be incorporated with the data and object types.

3.4.2. Information Requirements for Annotation of Music by producers

In order to identify the information required by music producers for music annotation ontology we have focussed on the process of creation as well search and retrieval of music metadata. Music consumers are only interested in typing a few keywords to find the intended piece(s) of music from a minimum number of search results. To maximize the possibility of getting satisfactory search results they require information about how to create description of musical objects and the representation of music data created by them may be manipulated to yield the best matched search results. In figure 3.5 I intend to show the connection between the process of music data creation and retrieval. The contribution under this thesis will attempt to provide the music producers with the required knowledge and information about how they can create description of musical objects, subjects and their inter-relationships. The music annotation ontology (***mpeg-7Music***) will uniquely focus on providing concepts and properties that will associate objective metadata that corresponds to the mapping of MPEG-7 timbre and melody descriptors data with subjective musical concepts as perceived by music producers during music data creation process. Subjective metadata will correspond to the proposed vocabulary of structured metadata that are being modelled now from the following discussion on figure 3.6 and 3.7.

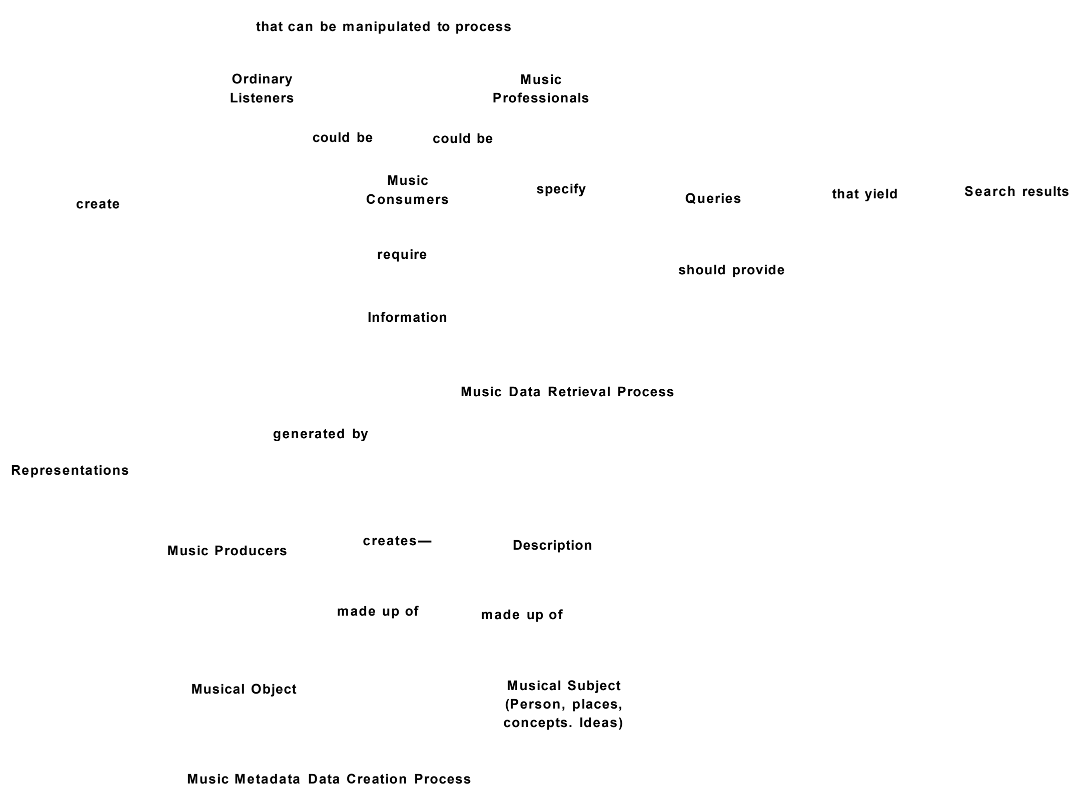


Figure 3.6: Utilization of Music Annotation Ontology

Let's have a look at how the proposed music annotation ontology will be utilized by music producers, specifically for the purpose of music metadata data creation. As shown in figure 3.6, music producers require information to annotate a music file that ***is described by*** the objective content (MPEG-7 elements) as well as it ***reflects*** the subjective content in it. These two types of content information - both objective and subjective content of the music file are explicitly defined in the proposed ontology (***mpeg-7Music)*** and are joined together by data type and object properties. Subjective content is also associated with contextual information that ***includes*** Time. The most important contribution of this ontology is that it ***includes*** Musical Concepts ***from*** Timbral and Melodic Descriptors those of which in turn ***characterized by*** concepts created specially to ***denote*** musical concepts under this ontology. The vocabulary chosen to denote the musical concepts that have been considered to populate the list of musical concepts vocabulary were selected from a survey conducted at MIT media lab with a view to collect phrases used by ordinary listeners (Sarkar et al., 2007) and online music teaching module(Schmidt-Jones, 2010).

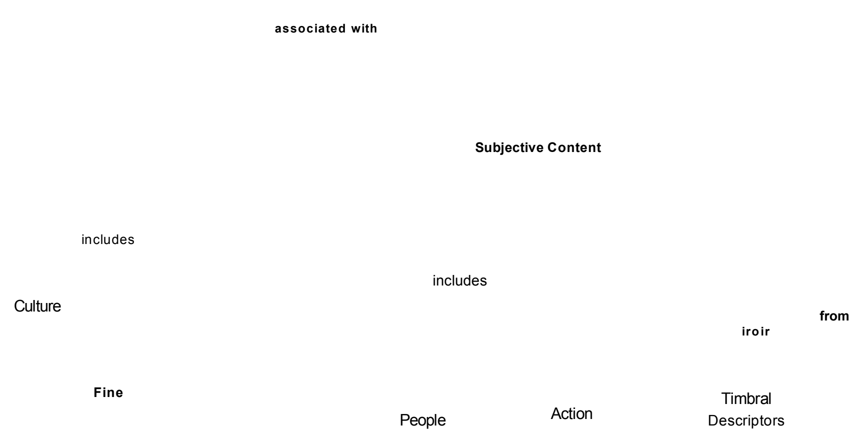


Figure 3.7: Use Scenario of the Music Annotation Ontology

Figure 3.7 shows that users annotate their music files with phrases associating with timbral descriptors as well as melodic descriptors but they do not want/need to worry about the objective content of music's acoustic data. The ***mpeg-7Music*** ontology will automatically associate objective data with musical concepts and the user will be able to select from the built in phrases (or able to add more phrases) to annotate music files. Automatic association of the objective content will be provided through few built in rules in the ontology. The design plan of the proposed ontology was not to constrain users with those rules only, users will be able to use those rules to create new rules of their choice to incorporate their specific interest (e.g. cultural) – addition of more rules will be considered as a future direction to improve the ***mpeg-7Music*** ontology.

While modelling music annotation ontology the following simple words have been selected as music consumers' phrases to support for subjective content. These words have been chosen from a previous research (Sarkar et al., 2007) and online music module (Schmidt-Jones, 2010). The figure 3.8 shows below a snapshot of those phrases.

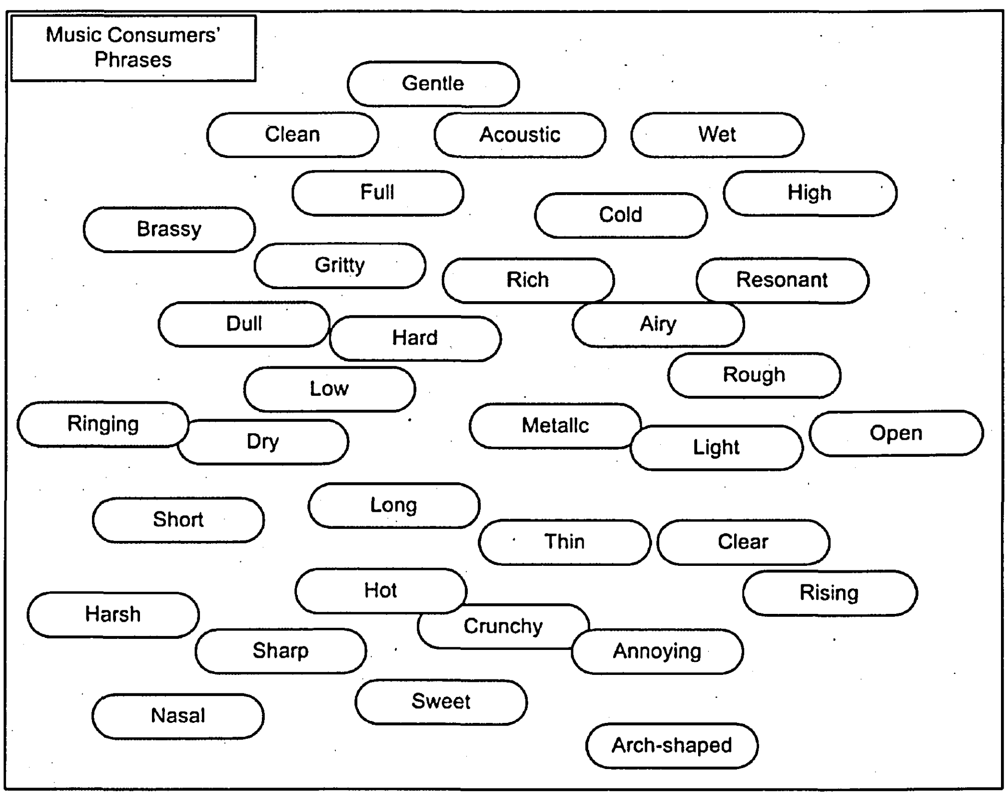


Figure 3.8: End Users' Phrases for Musical Concepts of the Music Annotation Ontology

The phrases compiled by the survey conducted at MIT Media Lab (Sarkar et al., 2007) describing timbral features and melodic features by Schmidt-Jones (2010) as shown in figure 3.8 are unstructured and are very subjective - because a single phrase may be used to derive multiple meaning by different music users. Besides, such phrases are quite ambiguous in meaning as may be understood by music consumers to describe a particular music item and associate these phrases with different musical dimensions. As these phrases have not been organized into any hierarchy and structure these are not suitable to be used for categorization or annotation of musical objects. As a result it requires us organizing these phrases into a conceptual structure for further utilization of these compiled phrases in facilitating the task of music annotation. The next section will discuss in detail how some of these phrases have been grouped together to cover timbral and melodic concepts in relation to MPEG-7 acoustic descriptors that will form the basis of the proposed contribution of this thesis.

3.4.3. Modelling MPEG-7 Low Level Audio Descriptors

The domain of the proposed ***mpeg-7Music*** ontology involves semantic descriptions of music files as understood by both music producers and consumers because the targeted users of the proposed semantic annotation of music items are the producers while the generated semantic annotations will facilitate the music consumers to search and retrieve music items. It is quite difficult for ordinary consumers to describe music effectively using terms like timbre, melody etc. as they are often used to represent complex musical dimensions. Even for the music producers it is difficult to identify the relationship between MPEG-7 objective contents (from MPEG 7 Part 4) and the consumers' subjective description. So, the ***mpeg-7Music*** ontology incorporated MPEG-7 descriptors mapped to its data type properties liberating the naive music users from the complexity of meaning of MPEG-7 descriptors.

The proposed music annotation ontology has generated appropriate concepts and properties to represent different timbral and melodic class concepts that are aimed to enable the users to derive implicit association with MPEG-7 low level descriptors. This required extracting MPEG-7 features to be linked to timbral and melodic descriptors. The next section studies existing available tools that enable to generate MPEG-7 features output from music files.

3.4.4. Modelling Principles and Steps

This section describes the methodology that has been applied for the definition of OWL ontology for annotating music files; that fully captures the MPEG-7 Audio parts. The principles followed in this thesis were:

Ontology Reuse: The ABC multimedia ontology is considered as upper ontology reference for developing MPEG-7 Audio annotation ontology. The proposed ontology either has reused or extended upper ontology concepts or properties.

Creating new concepts: To capture the semantics MPEG-7 Audio part the proposed ontology has created new ontology classes and properties.

Standardization of the existing music ontology: Existing OWL encoded music ontology concepts and properties have been taken into consideration that creates a further direction on how to lift widely accepted vocabularies with dominant standards such as MPEG-7.

Use of open source tools: The Protege OWL editor has been used to develop and validate the ontology.

Based on these principles, a two step process has been followed while developing the ontology:

Step 1-Extending Upper ontology concepts/properties and Utilizing existing widely accepted vocabularies: Modelling of the proposed ontology has started by extending the of ABC upper ontology's time concept that represents date/time points and duration in order to incorporate two concepts pertinent to musical time defined by Music ontology. Music ontology timeline concepts are based on physical/universal timeline and relative timeline or duration/interval to associate editorial metadata and media related metadata respectively. But, ABC ontology's time concept does not consider these two perspectives of time related metadata. Inspired by these two timeline concepts, I proceed by extending ABC's time concept with two class concepts namely PhysicalTime and MediaTime. To address temporal properties for each of these time subclasses the proposed ontology has created two properties onPhysicalTime and onMediaTime where each time point was identified by a xsd:dateTime and a xsd:duration respectively.

Step 2-Creating New concepts/properties: MPEG-7 timbre type relies on the estimation of the fundamental frequency and the harmonic peaks of the spectrum or on the temporal signal envelope. To incorporate MPEG-7 timbral descriptors mentioned above, timbralDescriptor datatype property has been created in the ***mpeg-7Music*** ontology and timbral descriptors are added as sub-types under this data type property. These data type property definitions create provision for lifting timbral descriptors values from MPEG-7 XML output.

To enable music producers to describe a music segment with simple words (to designate timbral characteristics) the annotation ontology comprises MusicalConcepts class's subclass TimbralExpression as well as the *timbralDescriptor* data type properties that are generated from a set of vocabulary based on (Sarkar et al. 2007).

MPEG-7 Melodic Descriptor and Description Schemes (DS)–MPEG-7 low-level descriptor Audio Fundamental frequency type (AFT) is a good predictor of musical pitch and it may be possible to make meaningful comparisons between data labelled with a melody descriptor, and data labelled with fundamental frequency. To aid further search based on melody information *melodyContourDescriptor* datatype has been created based on pitch interval and ContourType. The ContourType in turn is dependent on AFFT and BeatType value.

The concept MelodicExpressions class has been defined as super class of categories (e.g. shape, motion) of using several individual phrases (e.g. rising, leap etc.).

MelodicDescriptos takes values from melodyContourType which is related to AudioFundamentalFrequency. The AudioFundamentalFrequency descriptor holds the frequency information in an AudioLLDScalarType, which can be a Scalar or a SeriesOfScalar type (Batke et al. 2004). Melodic expressions may be derived (Schmidt-Jones, 2010) verbally from shape of melody described by melody contour descriptor values, e.g. rising melody (for ascending notes denoted by contour value 1 & 2), arch shaped melody (in which melody rises and falls and again slowly) etc. Arch shaped melodies are easy to understand by ordinary users and may be described by contour value -1, 0 and -1.

Now, we will pay attention on the detail content of the ***mpeg-7Music*** ontology based on the steps we followed in developing the ontology. The next section provides the specifics of the ***mpeg-7Music*** ontology that have been developed.

3.4.5. The Content of the Proposed Ontology

The created ontology (***mpeg-7Music***) extends ABC ontology's AudioSegment concept by adding MusicalSegment concept and relationships representing the semantics of MPEG-7 audio features. It also formed inheritance relationship with Music Ontology (Raimond et al., 2007) by importing TimeLine concept and creates MediaTime subclass to Timeline that actually linked to MPEG-7 basicTimePoint and basicTimeDuration data types. Besides, it uniquely added MusicalConcepts class to cater for Timbral and Melodic Expressions as understood by ordinary listeners and these expressions had got defined association with MPEG-7 scalar and vector data types by appropriate data and object type properties.

Referring to section 2.2.3, I would like to emphasize here that the designed ***mpeg-7Music*** ontology is different from simply a controlled vocabulary for annotating music. Usually, ontology is conceived as a set of concepts, properties, instances and inference rules. It is true that ontology is a specialized category of controlled vocabulary. There are two views in the contemporary on what makes a controlled vocabulary to qualify as ontology. Graybeal and Alexander (2009) assert that first one only concentrates on representing the controlled vocabulary using ontology encoding languages e.g. OWL and further subtleties of classification are not considered important. Such an approach focuses more on the machine process-able semantics of the vocabulary and interoperability over different systems but does not consider its semantic richness. The other view puts more attention on the concepts and requires that concepts are defined explicitly and at least some of them are defined as classes and the ontology must conform to strict hierarchical subclass relationships between the classes. As a result, such an approach puts more attention on defining rich semantic metadata and this can also be encoded using ontology encoding language such as OWL. In this approach, the strict relationship between classes through definition of appropriate properties– is designated as ontology that can include (but is not required to include them all) classes (types of things), instances (individual things), relationships among things, properties of the things as well as functions, constraints and rules related to the things depending on from which perspective they are represented. According to the second approach this thesis considers creating a semantically rich structured vocabulary that can be represented using OWL. The ***mpeg-7Music*** annotation ontology qualifies to be lightweight concept ontology to enable music producers to annotate any music segment from both the view mentioned above as it is represented using description logic based OWL species namely OWL-DL.

The ***mpeg-7Music*** i.e. the music Annotation Ontology developed under this thesis acts as a vocabulary that captures MPEG-7 audio features and music consumers' keywords to describe music or music segments' subjective contents. The ***mpeg-7Music*** ontology contains components of OWL ontology i.e. classes, properties and individuals to interconnect concepts from three domains – a subset from MPEG-7 vocabulary, musicological concepts and end-users tags. The ***mpeg-7Music*** ontology presented here contains a class hierarchy showing a strict subclass structure of categories (classess); where individual keywords defined under each class may be related with properties defined in its property hierarchy. Next, the detail of different elements (classes, properties and individuals respectively) defined in the ***mpeg-7Music*** ontology will be presented in a top down manner for clarity though initially it was modelled it in a bottom-up fashion. Figure 3.9-3.11 depicts the content of the ***mpeg-7Music*** ontolgy.

Class definition and Class Hierarcy: Classes provide concrete representation of concepts and may be interpreted as sets or categories that contain individuals. The initial Class Hierarchy as shown in figure 3.10 contains a subclasses derived from owl:Thing class. The ResourceFormat class served to facilitate the annotation process by defining subclasses with audio file formats that organized the resource and was intended for checking if the file to be annotated contained audio material. MusicalConcepts class formed the top level concept to create a category of different (music consumers') tags. Musical segment class was created as an extension to MPEG-7 AudioSegment that was used to create association of user tags with the musical resources. The TimeLine class was designed to identify physical and media time based annotations.

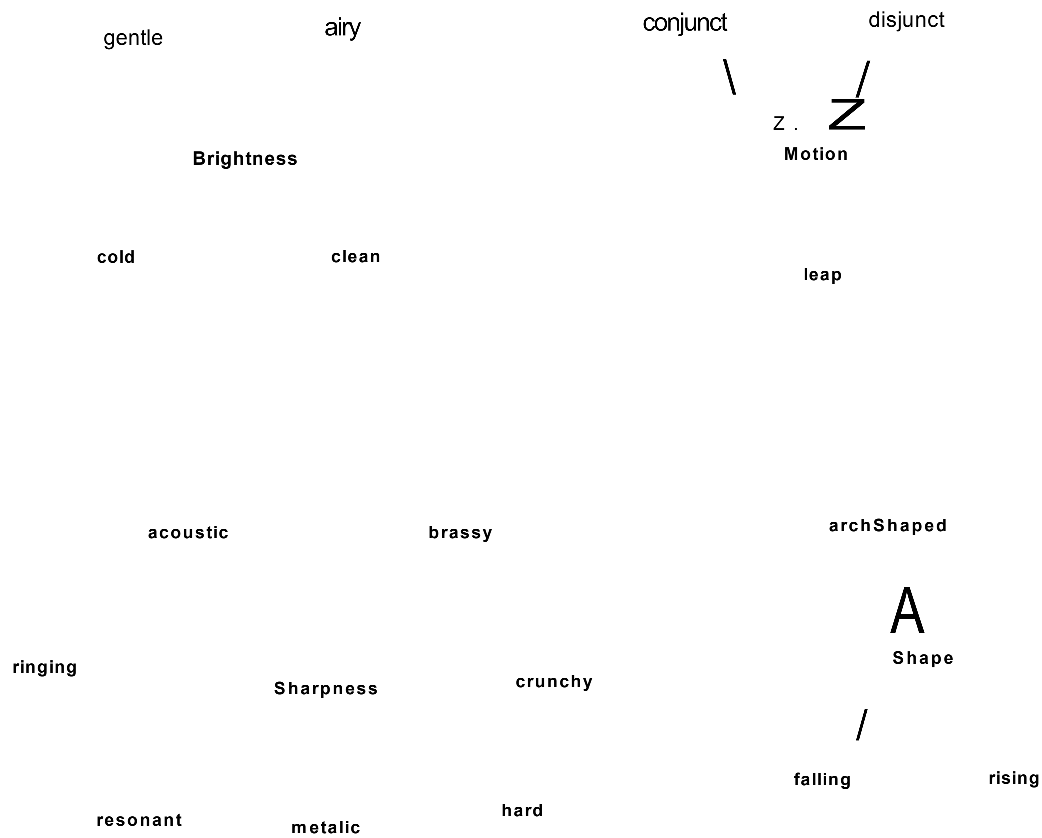


Figure3.9: Instantiation of Individuals

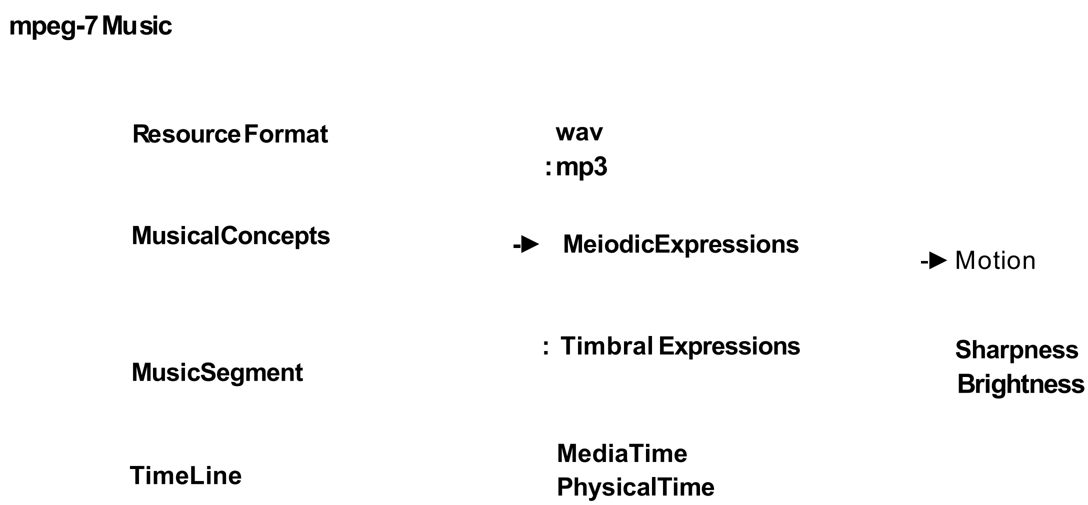


Figure3.10: Class Hierarchy

Determination of Properties and property Characteristics: Properties (also known as relations) are binary relations that link two individuals. E.g. 'describeBy' property was created to link instances of MusicalSegment and MusicalExpression classes. The two main types of properties of ***mpeg-7Music*** ontology were- object properties and datatype properties. Object properties links individuals to individuals and data type properties links individuals to XML datatype. The object properties of the mpeg-***7Music*** ontology were designed to link individuals from music consumers' tags to individuals from musicological concepts. E.g. *relatesTo* property was created to link timbral tags (instances of timbral expression class) with instances from musicological concepts classes and subclasses (e.g. sharpness, loudness or brightness). Example of datatype property included mpeg7DataTypeDescriptor: melodyContourValue that linked MusicalSegment class instances with XSD:datatype extracted from MPEG-7 audio encoder output.

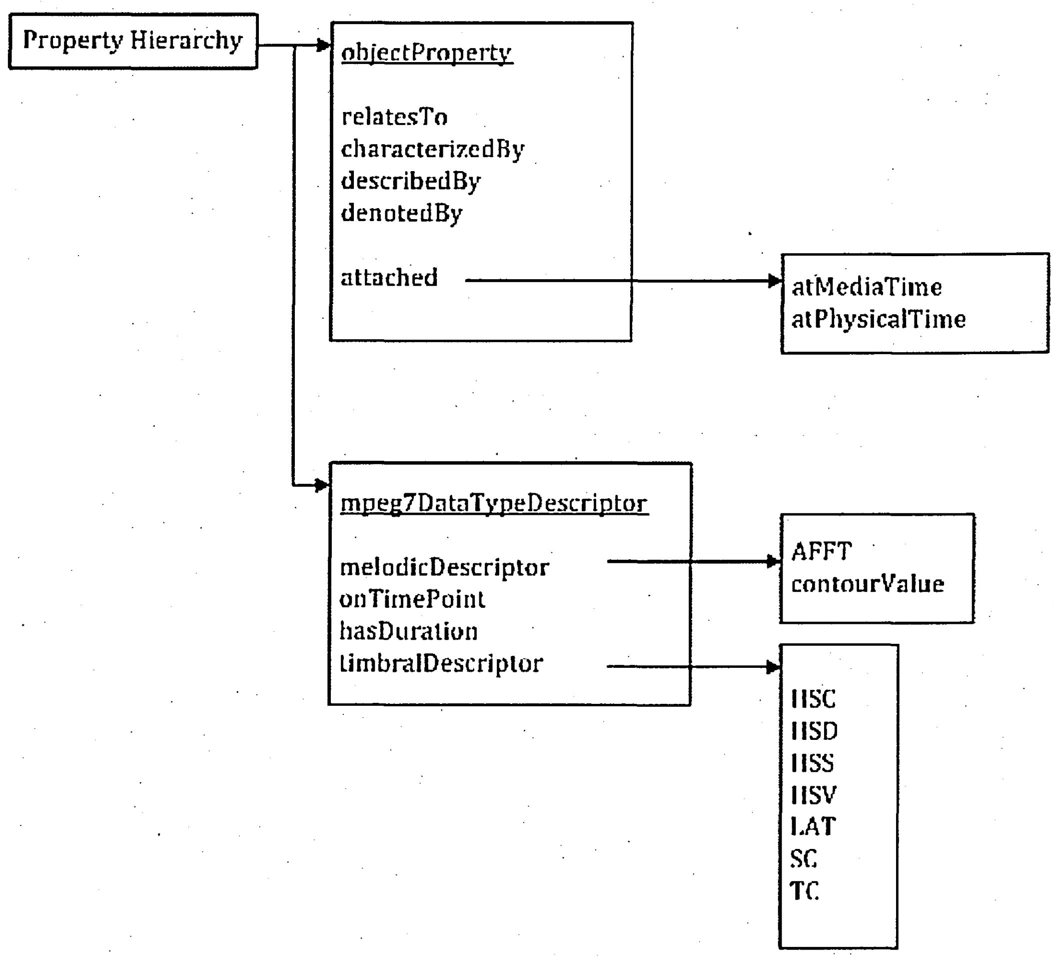


Figure 3.11: Property Hierarchy

The property hierarchy (in figure 3.11) shows both data type properties and object properties defined to associate MusicSegment and MusicalConcept (it's subtypes and their individuals) as defined by the class hierarchy (figure 3.10).

characterizedBy is a symmetric property. It describes: MusicalSegments *characterizedBy* MusicalConcepts and vice versa. It also linked ResourceFormat with MusicalSegments and thus this property has got transitive characteristics as it supports assertion like: ResourceFormat *characterizedBy* MusicalSegments; MusicalSegments *characterizedBy* MusicalConcepts.

relatesTo is a transitive property ; it linked MusicalSegments with ResourceFormat and MusicalConcepts. Provision was created for the music producers to choose how they feel it appropriate to describe their music item with two object properties (characterizedBy and relatesTo) having different characteristics.

denotedBy links MusicalSegments with TimeLine and its subclasses. A particular music segment can have at most one MediaTime and PhysicalTime instance. So, denotedBy is a functional property.

describedBy links MusicalSegments with MusicalConcepts. No restriction was applied on this property as the same MusicalSegment may be described using multiple musical concepts.

attached is an object property that links MusicSegment with TimeLine. Subproperties defined under attached property are *atMediaTime* and *atPhysicalTime*. These two sub-properties connected musical segment with instance of TimeLine and its subclasses MediaTime and PhysicalTime instances. Instances of TimeLine class may take values from two data type properties hasDuration (MPEG-7 xml datatype basicDuration) and onTimePoint (MPEG-7 xml datatype basicTimePoint).

timbralDescriptor and *melodicDescriptor* data type property contains a collection of MPEG-7 XML datatype and they describe MusicalSegment. The MPEG-7 Audio Encoder tool (as mentioned in section 3.1.3) was used to extract the low level audio features for each of the music file. However, the MPEG-7 standard and the ABC ontology did not contain classes and properties to support representation of music content by music producers with audio features as described in section 3.2.2. Therefore, ***mpeg-7Music*** extended ABC multimedia ontology by defining subclasses of AudioSegment class to music features and constructed properties both subjective (that relates both melodic and timbral descriptors with musical expressions chosen by music producers) and datatype (that attaches MPEG-7 low level descriptor values with Music Segments). Figure 3.11 shows a summary representation of the constructed properties.

Many of the MPEG-7 low level audio feature related concepts are too specific to be comprehensible to anyone but acoustic experts. It is not reasonable to expect ordinary music lovers (as well as musicians) to understand what 'Audiofundamental Frequency' or what values the property 'mediaTime' may take in the units of milliseconds. Instead of these specialist terms, the music lovers use more commonly known terms like 'sweet', 'short' and 'happy' to describe music tune characteristics. Figure 3.9 shows the categories of such common phrases that were used to model as instances under musical concept subclasses.

Creation of Individuals: The word individuals denotes instances of classes i.e. objects from the domains e.g. *melodyContourDescriptor*, *spectralCentroid* value from MPEG-7 vocabulary; musicological concepts e.g. melodic concepts (rising, falling etc.); ordinary users' tags (e.g. timbral tags – crunchy, resonant etc.). Subclassess defined under MusicalConcept contain categories for melodic and timbral tags namely MelodicExpressions ands TimbralExpressions respectively as shown in figure 3.10. Music consumers' tags were instantiated from those categories. To create the grouping of consumers' tag as subclasses under MelodicExpressions and TimbralExpressions classes, this thesis relied on MPEG-7 data content quality insights as used in (Schmidt-Jones, 2010) and (Mitrovic et al. 2007) respectively. Figure 3.9 shows the individuals that were created under each of the subclasses. For example, motion of a melody may be either termed as conjunct or disjunct or leap. Depending on whether a melody rises and falls slowly/quickly it is generally referred to as *conjunct* or *disjunct* repectively. If it is a mix of both then it is called *leap*.

Having defined the detailed ontology structure the rest of the problem was how it could be enabled to associate MPEG-7 descriptors mapping with MusicSegment class during annotation by music producers. To solve that few basic rules were defined to translate low-level acoustic terms gathered from MPEG-7 description such as 'spectrum centroid' (SC) to more familiar commonly-used intermediate-level terms like 'long' and 'short'. Donnadieu (2007) and Sarkar et al. (2007) researched such links between low-level features and commonsense terms., They found, for example, that humans consider a piece of music characterizing dull or bright based on perceptual dimension linking to Spectral centroid (SC) - this fact was used to construct the following association rule to link timbralDescriptor:SC with MusicalSegments.

Timbral Rules
Rule 1-BrightnessFeature Association Rule: ***mpeg-7Music:***MusicSegment(?x) □ ***mpeg-7Music:*** TimbralDescriptor(SC) → ***mpeg-7Music:***characterizedBy(?x, Brightness)
Rule 2- SharpnessFeature Association Rule: ***mpeg-7Music:***MusicSegment(?x) □ ***mpeg-7Music:***TimbralDescriptor(HSC) □ ***mpeg-7Music:***TimbralDescriptor(HSS) → ***mpeg-7Music:***characterizedBy(?x, Sharpness)

The above rules associate different MPEG-7 data descriptors (those were modelled as data type properties in ***mpeg-7Music*** *ontology*) with MusicSegments instances using the characterizedBy object property (which is actually a symmetric property).

The TimbralExpression class in the ***mpeg-7Music*** ontology contains two broad categories of music consumers' phrases named as Sharpness and Brightness (as shown in figure 3.9) and each of these sub-categories contains the instances of individual phrases borrowed from (Sarkar et al., 2007) that only provide a set of key phrases but ***mpeg-7Music*** was added with category hierarchy for these phrases and the timbral tags that were used for the canonical rules (Rule 1 and Rule2) according to research efforts of Agostini et al. (2003) and Misdariis et al. (2010). The music producers can use this rule to annotate any music segment with TimbralExpression phrases or create more rules if they feel necessary. If they want to mark the music segment to have bright timbre (using instance from TimbralExpression class) then they can easily do that using Rule 1 and 2 (mentioned above); these rules were created to form the abstraction layer for associating spectral centroid datatype value of the underlying audio.

To generate these rules, SWRL (Semantic Web Rule Language) notations (Horrocks et al., 2003) were used to specify rules from the proposed ontology classes, properties and individuals expressed in OWL. The OWL notations provide constructs and syntax to represent concepts, properties and individuals that were mentioned earlier in this section. In order to specify rules concerning these concepts, properties and individuals additional notations were needed e.g. *'and', 'or', 'implies'* etc. SWRL was chosen because it is available as a Protege OWL plugin and Protege tool was used to encode the ***mpeg-7Music*** ontology. SWRL could be used within Protege as the SWRL Tab (Connor et al., 2005) and it permits the creation and execution of SWRL rules and a variety of third-party rule engines (e.g. Jess rule engine) may be plugged in to add inference. The SWRL API (Connor, nd) provides a mechanism to create and manipulate SWRL rules in an OWL knowledge base. SWRL is based on RuleML that aims to provide a shareable, XML-based rule markup language (Boley et al. 2001). It provides format that ensures that rules are machine readable and interoperable with existing tools and standards. SWRL combines RuleML and OWL (Horrocks et al. 2004). The novel contribution of this thesis here is the inclusion of numerous built-in relations that creates the connection between low level audio terms with intermediate level terms to facilitate user annotation. A typical example of a rule in the problem domain that this thesis is addressing is the rule below which is applied to recognise a song from its melodicCharacteristics that associates the musical resource with the fundamental frequency of the underlying audio.

Melodic Rules
Rule 3- MotionFeature Association Rule: ***mpeg-7Music:***MusicSegment(?x) ▯ ***mpeg-7Music:***MelodicDescriptor(AFFT) → ***mpeg-7Music:***characterizedBy(?x, Motion)
Rule 4- ShapeFeature Association Rule: ***mpeg-7Music:***MusicSegment(?x) ▯ ***mpeg-7Music:***MelodicDescriptor(contourValue) → ***mpeg-7Music:***characterizedBy(?x, Shape)

In section 3.3.8, initially I mentioned the intention to use Protege 4.0 (that supports OWL 1.1) to develop the ontology. In order to work with SWRL, SWRLTab (Connor, nd) is need but it is still under development to be ported to Protege 4.0. So, it was decided to represent the rules in SWRL using Protege3.4.1 but this version of Protege does not support OWL 1.1. For ease of implementation, only object property characteristics were defined (disjoint role characteristics requires OWL 1.1, so were not modelled) and generated SWRL rules covering OWL 1.0 features only. Further rules may be generated using SWRL built-ins such as lessThan, equals etc.

Now, the details of the ontology - ***mpeg7Music*** outlined above was intended for annotating digital music. To illustrate how ***mpeg7Music*** may be used as a backbone for annotation process of digital music, a customized semantic annotation tool was designed and implemented. The next section provides a brief overview of the tool and presents examples of how annotations can be generated using the main contribution in this thesis that is the ***mpeg7Music*** ontology.

3.4.6. The Simple Semantic Annotation Tool

Having designed the music annotation ontology enriched with MPEG-7 features; modelled from a set of music consumers' tags, the remaining task was demonstrating the scope of this ontology to enable music producers for annotating music. This led to look back to the research field of Web 2.0 (Zhang, 2007) technologies that present applications for developing folksonomies. It is true that folksonomies did not provide techniques to develop meaningful annotations but folksonomies can be utilized as a means for cheap and quick alternative for automatic annotation but the created annotations are not designed for machine processing (Bateman et al., 2006). In the design of the proposed ***mpeg-7Music*** annotation ontology the idea of using machine process-able encoding powered Semantic Web technology's OWL standard was considered. Though Web 2.0 applications appear to be an excellent opportunity to create semantic resource, in the field of Semantic Web technology it is still suffering from few shortcomings in the context of developing semantic web applications. Initially, annotations generated using Web2.0 applications were not aimed for machine processing and hence their automatic manipulation narrows down to the complexity of natural language processing (Gruber, 2008). Besides, looking at annotation tools' user intervention requirements it is usually found that manual annotation tools do not seem to be effective because of error-prone annotations and the laborious effort needed to perform the annotation task makes it an unattractive option. On the other hand, fully automatic annotation tools rely mainly on huge training data needed for effective machine learning algorithms and such tools are unable to capture the high-level meaning conceived by human users. So, another limitation of Web2.0 applications to be adapted for a semantic annotator application is to overcome 'semantic knowledge acquisition bottleneck' in terms of how user intervention may be incorporated in the annotation process and to what extent automation may be enabled. Consequenently, in order to effectively demonstrate the suitability of the proposed ontology a semi-automatic annotation tool was designed as a trade-off to bring the best of both worlds of Web2.0 and Semantic technologies (Damme et al., 2007).

Section 2.2.4 outlined the need for a customized semantic annotation tool that satisfies the following three requirements:

- Ontology based annotation for achieving interoperability
- Standard representation of annotations
- Type of content of heterogeneous multimedia resources

From the survey of existing annotation tools it was found that for annotating digital music currently available annotation tools did not provide adequate support for music consumers to be able to semi-automatically annotate music items of their choice as well as satisfy all three of the above mentioned requirements. So, the simple semantic annotation tool was designed to show the applicability of the ***mpeg7Music*** ontology. Figure 3.12 shows the high level system view of the tool that uses ***mpeg7Music*** as backbone to guide the annotation process. MPEG-7 feature was extracted from the music file (to be annotated) automatically and mapped to the ontology through the properties defined in the ontology. Then the user was presented to associate higher level concepts with these auto generated features. So, a part of the annotation was automated and to complete the process of annotation it required user intervention. Hence the designed annotation tool was classified to be semi-automatic.

Annotation Ontology

Annotation Tool ► Annotations

Music File

* MPEG-7 Feature
Extractor output

Figure 3.12: High level System view of the Semantic Annotation Tool

The high level system view of the simple semantic annotation tool as shown figure 3.12 has two distinct phases of workflow: Lifting MPEG-7 metadata by associating it with annotation ontology (figure 3.13) and the actual annotation process with steps as detailed in figure 3.14.

Lifting MPEG-7 metadata: When a music file is loaded into the annotation tool several background steps are executed without any user intervention as displayed by figure 3.13. At first, MPEG-7 feature extraction output is generated automatically for that particular music file using ***MPEG-7 Feature Extraction Tool.*** A freeware MPEG-7 audio encoder tool (Crysandt, 2005) was utilized to generate MPEG-7 XML output from the music file. Secondly, MPEG-7 datatype values were extracted from that XML output using the custom built ***XML Data Source Processor.***

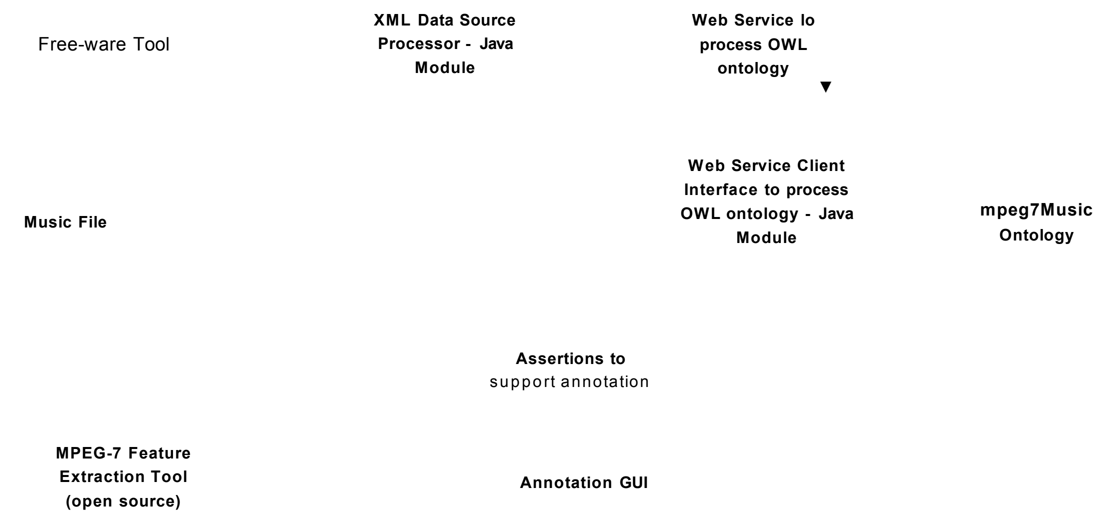


Figure 3.13: Workflowfor lifting MPEG-7 Metadata[16]

These data type values were then mapped to the ***mpeg7Music*** ontology using the melodic and timbral rules in the ontology. To perform the mapping process, a web service was designed to process OWL ontology using Protege OWL API17. The reason to use web services was to create interoperable service oriented architecture (Mahmoud, 2005) for the proposed annotation system. Having designed the web service (OWL Ontology Processor - Web Service) to fit the need, a web service client was also created to interface the annotation tool with the web service. The web service client was used to map the data types extracted from the XML data source processor with properties and concepts of the ontology. Association rules for linking MPEG-7 XML tags defined by the annotation ontology preserved the connection of the created annotations with MPEG-7 description of the underlying audio of music files. After the data type values were mapped then the assertions generated were made available for the graphical user interface step 3 with the help of the ***mpeg-7Music*** music annotation ontology.

16 The feature extraction tool in the red-shaded rectangle was the open-source free-ware from MPEG-7 Audio Encoder (2008). The XML outputs from this tool were processed by our Simple Semantic Annotation tool. Modules inside the orange shaded rectangle were designed and developed by us.

17 http://protege.stanford.edu/plugins/owl/api/

Once the generated assertions have been displayed to be used by users then the next part of the annotation workflow that relied on user intervention began the actual annotation process by music producers.

Figure 3.13 shows a part of the actual annotation process (i.e. at step 3) where the output of the automated lifting of MPEG-7 metadata was presented for user input to associate with ontological concepts. Now, each step of the actual annotation process will be discussed next as depicted by figure 3.14.

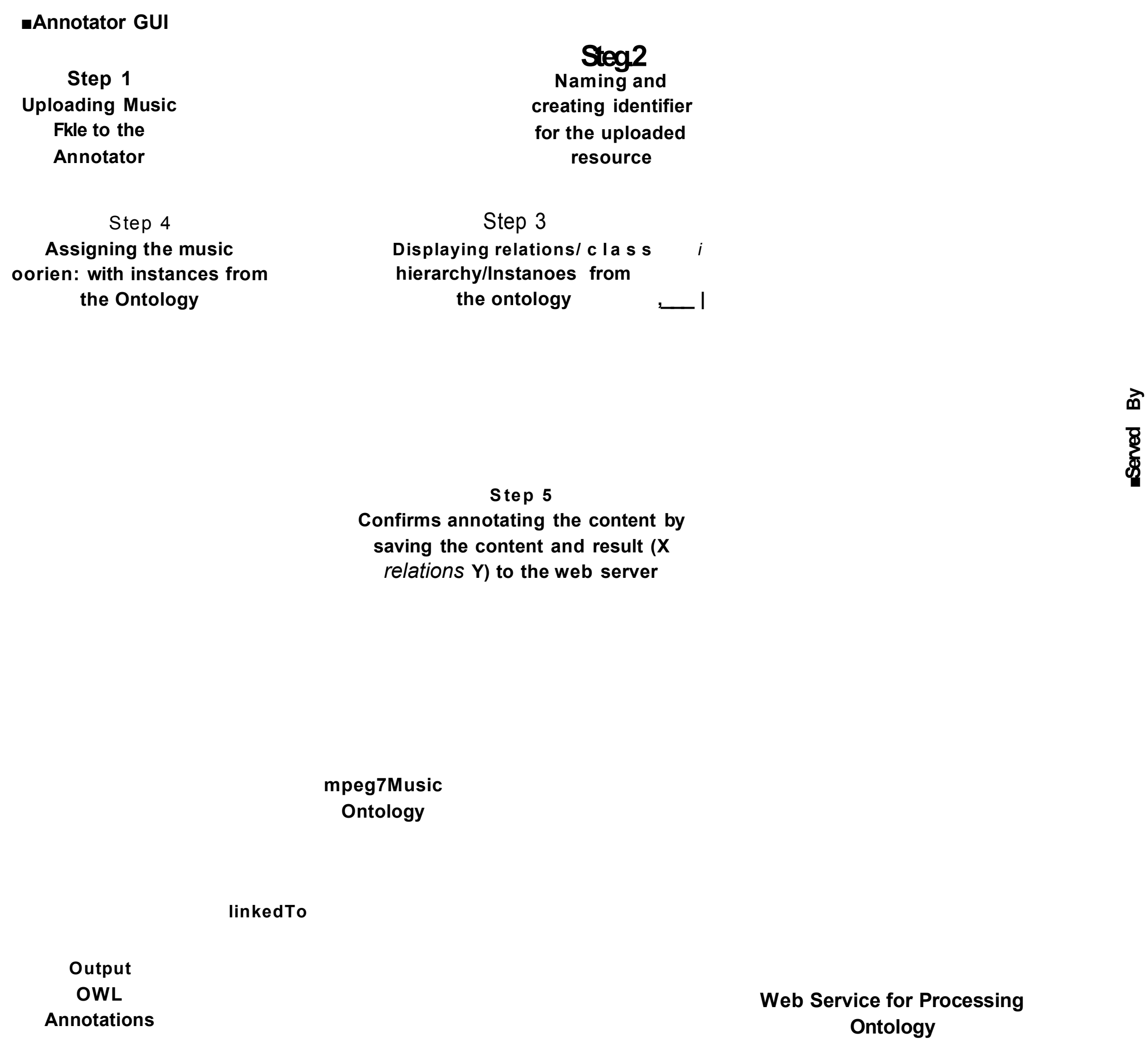


Figure 3.14: Semantic Annotator- Application Architecture

Actual Annotation Process: The actual annotation process of semantic annotation tool was based on a simple five steps process; those will now be explained with reference to figure 3.14.

I. StepI- ***Loading and naming the Resource***: When a user loads a resource (i.e. music file) using the annotator interface a Globally Unique Identifier[18] (GUID) is created in order to provide a unique reference number to the loaded resource and that reference is unique in any context. The loaded resource may either be uploaded to the annotation server or downloaded from any web location. In case of downloading the resource (to avoid any copyright restriction) the URL (Uniform Resource Locator) of the content is stored instead of the resource itself. For example, one can load the resource and create an identifier X to denote it.

II. ***StepI-Identifying the resource type:*** In order to associate semantic concepts to ***X***, the annotator application first determines the type of content (whether the content is mp3, wav or in real audio format) from the resource's file format with respect to the ResourceFormat class as defined in the ***mpeg7Music*** ontology (The ***mpeg7Music*** ontology contains a classification of the content type (ResourceFormat class) depending on the file formats). Initially, this thesis restricted itself to only few file formats because the MPEG-7 feature extraction tool that was used to encode MPEG-7 XML output only supported for the defined types under ResourceFormat class in ***mpeg7Music*** ontology (figure 3.10).

III. Step ***3-Displaying relations:*** In this step, concepts and individuals from the domain ontology is displayed for the user to associate with the resource's content through the relation that the user has chosen in step 3.

[18] http://en.wikipedia.org/wiki/Globallv Unique Identifier

IV. Step 4-Assigning MusicalConcepts from the instances of the annotation ontology: Users may click on the individual of choice to relate it with the resource. Following the example of resource X, a user may create annotation in this step as follows:

"**X** *relatesTo* (TimbralExpression-Sharpness) ***acoustic***

After creating a single annotation the user goes back to step 3 where s/he can choose another relation/property to associate concepts from the property hierarchy of the ontology displayed in step 3 again otherwise if the user chooses to finish the process of associating domain ontology concepts then s/he goes to step 5.

V. Step 5-*Confirmation of Annotation*: In step 4, the user may have a final check of the created annotations and confirm the results through saving/storing annotations functionality that stores the results of the annotation process and finally completes the annotation process.

Software Modules to support the workflow: Three software modules were designed and implemented as described below:

1) XML Data Source Processor: As mentioned in section 3.3.2, this contribution used JAVA API for XML processing. JDOM was used to parse MPEG-7 encoder output.

2) Web Service for processing (Protege OWL) ontology: The ***mpeg7Music*** annotation ontology was developed using Protege ontology editor and the ontology was encoded in OWL 1.0. To apply this ontology for enabling annotation process - parsing and displaying the class, property hierarchy; linking MPEG-7 datatypes with the object properties using the rules defined in the ontology– a web service was developed for processing OWL ontology using Java based Protege OWL API.

3) Semantic Annotator Graphical User Interface: The semantic annotator GUI was designed and developed in a way that it uses the annotation ontology as a backbone for annotation process that works using a five step process (as in figure 3.14). The interface and the associated functionalities were developed using JSP technologies.

The above mentioned software modules were created using NetBeans 5.5 Java IDE. To store the output of annotation process, MYSQL database tables were designed to save the created annotations that may be utilized by semantic search interface as we showed in figure 2.7 (section 2.1.2).

The Simple Semantic Annotator application introduced three innovations as specified below. The first two were concerned about introducing a balance between user intervention and automation. The third one is related to the design of the annotation procedure with MPEG-7 features information and structured metadata vocabulary from ***mpeg-7Music*** ontology. The detail of specific points on three aforementioned innovations is as follows:

1. **The minimal set of mandatory information** was defined including:

    a) Globally Unique Identifier for the annotated resource - URL of the annotated resource (provided as the first step in annotation procedure) – created automatically.

    b) Title (provided manually).

    c) A list of assertions containing MPEG-7 datatype values from the ontology rules – automatically provided.

    d) A hierarchy of the concepts (with individuals) from the ***mpeg-7Music*** ontology (from the hierarchy the user can manually select).

    e) The property hierarchy is presented to the user from where the user can choose manually.

2. **Introduction of automation** wherever possible. From above, it is evident that 2 out of 5 mandatory elements are provided in a full automatic manner. Specifically, presentation of the MPEG-7 features is done automatically. The last two mandatory information required user intervention to choose but the hierarchy of concepts and properties were made

available to choose by the user; these two are partially automatic. Finally, depending on the selected annotation concept, the system will attempt to automatically select the matching relation, in order to complete the annotation assertion. Only in case that the selection was ambiguous, the user will be prompted to select the annotation relation manually.

3. **The design of the annotation procedure** followed the current tagging standards encountered in the Web 2.0. The annotation assertions are not formulated by selecting first predicate, and then object (depending on the selection of the predicate). Instead, the list of available concepts is displayed as if they were simple tags (also display of concepts in a tag cloud is provided, in order to make the experience even more similar to the Web 2.0). After the selection of a concept, the user is prompted to specify the annotation relation that will be used as a predicate in the assertion, only if there is an ambiguity detected in the analysis of the domain and range specifications of the selected annotation relation.

An Illustrative Example of Creating Annotations: Now, an illustrative example will show and describe each of the steps of the annotation process. Figure 3.15 (a-e) shows the screen mock-ups from the simple semantic annotator application by demonstrating interactivity of each step. The original screen shots that were created from the application were too blurry to be viewed when printed. So, these screen shots presented here have been generated using Adobe FireWorks[19] for clarity.

19 http://www.adobe.com/products/fireworks/web/

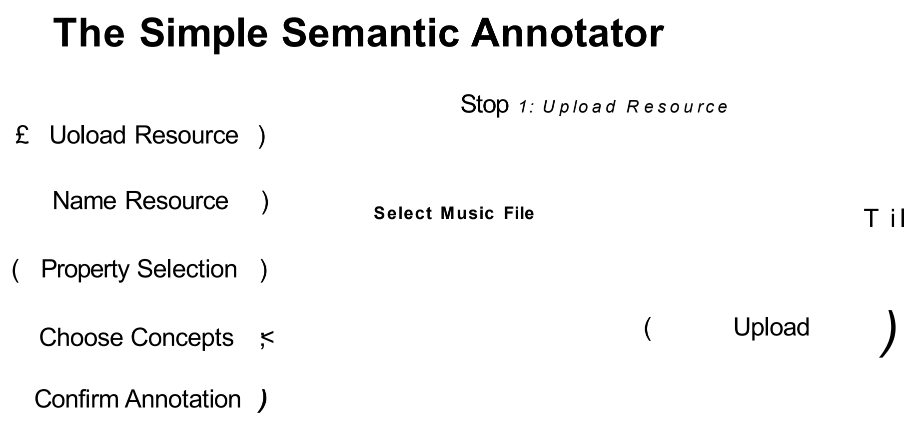


Figure 3.15(a): Simple Semantic Annotator- Step 1 GUI

Step 1 starts by providing browsing facility for the user to upload music file on the annotator tool. For example, the file that the user chooses is ***examplel.wav*** and the user clicks the 'upload' button.

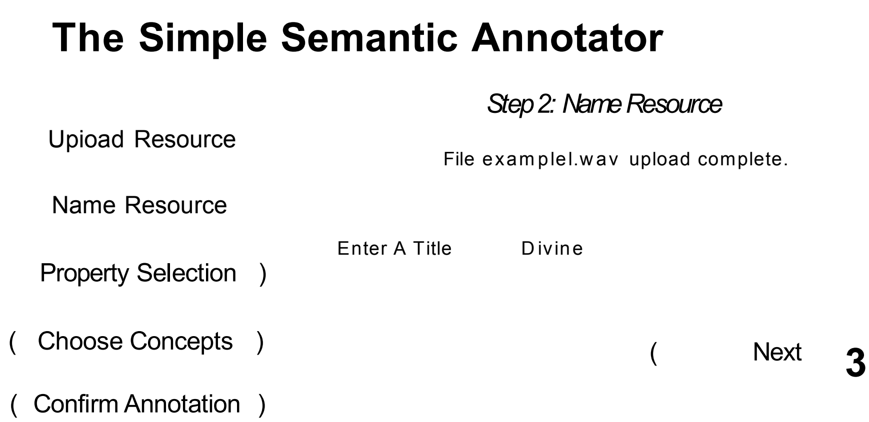


Figure 3.15(b): Simple Semantic Annotator- Step 2 GUI

Then Step 2 prompts the user to enter a title for the uploaded music resource. Here the title parameter is a mandatory input for annotation. Let the arbitrary title input be 'Divine' and the user clicks the 'next' button and before step 3 is displayed few process is executed at the back end. Firstly, the system then generates a globally unique identifier (GUID) for the resource so that the resource itself and the generated annotations can be linked together.

Secondly, the MPEG-7 feature extractor plug-in generates XML output (a sample XML has been attached in the Appendix C).Thirdly, the XML DataSource Processor uses the Timbral and Melodic rules to detect which tag from the MPEG-7 output to extract. For example, TimbralRule 1 associates SC (SpectralCentroid) parameter with the ***Brightness*** concept and MelodicRule 1 associates AFFT (AudioFundamentalFrequency) parameter with the Motion concept.

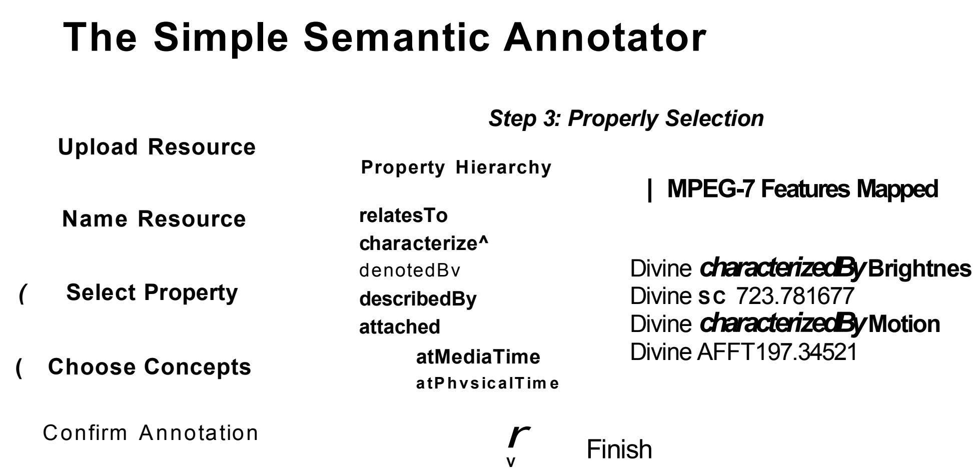


Figure 3.15(c): Simple Semantic Annotator- Step 3 GUI

As can be seen from figure 3.15(c) that step 3 of the GUI is divided into two panels - the left panel displays the property hierarchy for the user to choose for further annotation. The right panel displays a list of assertions auto generated from MPEG-7 output with reference to the rules defined in the ontology. Clearly, MPEG-7 assertions presented on the right panel is a static display, it's purposes are to provide the user suggesting on the concept to be chosen and the MPEG-7 data types associated with the resource Divine. The left panel displays the list of properties where the user can click on any of those object properties to associate individual concept of choice.

Let's consider if the user clicks on the ***relatesTo*** property then step 4 GUI is displayed.

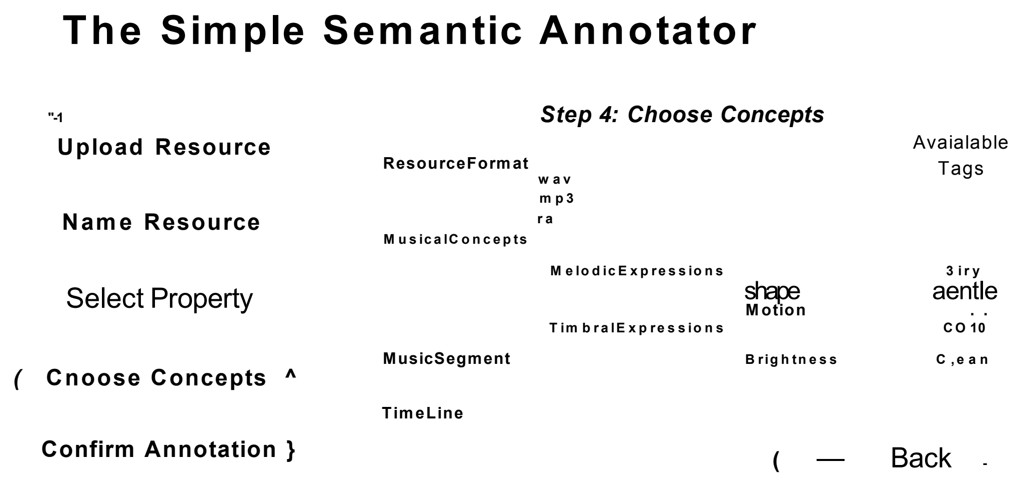


Figure 3.15(d): Simple Semantic Annotator- Step 4 GUI

Step4 of the GUI is also divided in two panels (figure 3.15(d)), the left one displays the class hierarchy from where the user can choose a concept and if a concept is chosen then the corresponding list of individuals are displayed on the right panel. Now, if the user does choose the Brightness concept then the individuals under motion sub-class is displayed as shown in figure 3.15(d). From the right panel then the user can select any individual and let's select **gentle**. Then to add more individuals the 'Back' button must be clicked to go back to Step3 and select any object property to come back again to step4 and so on until the user is happy. Let's assume the user chooses 'describedBy' property and arrives at step 4 and clicked on the Motion concept and selected conjunct individual from the right panel at step4.

After adding two individuals if the user goes back to step 3 and clicks the 'Finish' button then s/he arrives at step 5.

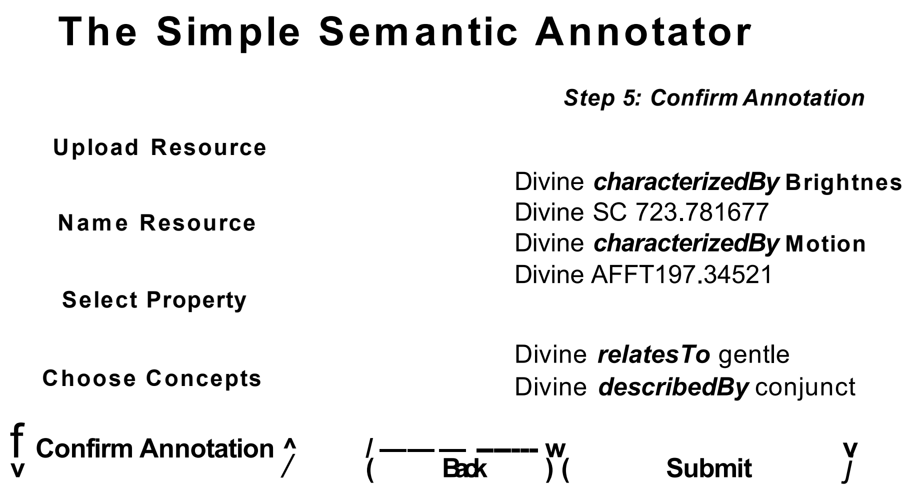


Figure 3.15(e): Simple Semantic Annotator- Step 5 GUI

The step 5 displays both the auto generated (maroon rectangle) and user created (light-green rectangle) annotations as shown in figure 3.15(e). If s/he is satisfied then click of the Submit button will save the annotations in the database. Otherwise the user clicks the 'Back' button and adds more or deletes any that was added before.

3.4.7. Summary

This section has presented the detailed design rationale for proposed music annotation ontology. It started from the famous Music Ontology by Raimond et al. (2007) in an attempt to create a scoping of the annotation ontology with music tagging/annotation by music producers. The ***mpeg-7Music*** annotation ontology re-used the time line concept of music ontology and then extended ABC ontology's AudiSegment concept by creating highest level concept 'MusicSegement'. This MusicSegement may be described by TimeLine concept or its subclasses. Besides, two other datatype properties ***(TimeInstant*** and ***TimeInterval)*** were created to incorporate MPEG-7 basicTimePoint and basicTimeDuration values that actually determined a MusicSegment using low level physical parameters. Besides, MusicSegment is composed of objective and subjective contents. Values associated with objective contents are collected from MPEG-7 low-level descriptors while subjective contents are collected from MusicalConcepts class and its subclasses (Timbral and Melodic expressions). Both Timbral and Melodic expressions were designed to take their datatype property values from MPEG 7 Audio high level tools structure. The High level tools structure was defined under the definition of Timbral and Melodic Expression class definitions.

The ***mpeg7Music*** ontology was coded using the Ontology Web Language (OWL). The ***mpeg7Music*** ontology was composed of classes and properties. Primitive Concepts were organized in a class hierarchy where each concept in the ***mpeg-7Music*** ontology being subsumed by ABC ontology's AudioSegment class. Each relation was defined by its domain and range.

For further utilization of annotated music files, some insights were shown in section 2.1.2 about how the created rules may be used by search and retrieval applications. However, while the ontology was tested and evaluated to assist in guiding the annotation process it was not intended to constrain music producers to the basic rules created in ***mpeg7Music*** ontology. More rules may be created using the SWRL utility (Connor et al., 2005). By enabling the music producers to create association among the MPEG-7 descriptors and musical concepts/phrases the semantic annotation tool actually engaged them to create/ populate the knowledgebase that may serve as a space for semantic search; where the search space is encoded using dominant semantic web standard (i.e. OWL 1.0) and the actual content description complies with industry standard for describing audio resources. Finally, the applicability of the ***mpeg7Music*** ontology was demonstrated to guide semantic annotation of digital music item using the simple semantic annotation tool. The contribution in thesis comprises of the ***mpeg-7Music*** annotation ontology and the simple semantic annotation tool as a demonstration platform that shows how music items may be annotated with MPEG-7 features.

4- Evaluating the Music Annotation Ontology

.1. Introduction

The contribution of this thesis is an ontology that was designed to facilitate the task of annotation of digital music by music producers. The need for such an ontology emanated from the limitations of current web search techniques to enable semantic search and retrieval of music of choice by ordinary listeners/ music consumers. From the literature survey it was found that searching of musical objects can greatly be improved by the use and generation of semantic metadata; but the generation of semantic metadata relies on at least two factors - quality of the semantically enriched metadata and automation of the process of semantic annotation and in the case of digital music we found that there is a huge community of music producers/ consumers who are not supported by efficient tools to be able to annotate music of their choice with semantic metadata. Besides, to be able to generate semantic metadata that is interoperable and exchangeable among the heterogeneous systems and networks, the representation of such metadata must conform to widely accepted industry standards both by multimedia data representation community standard(e.g. MPEG-7 Audio) as well as semantic metadata encoding standard (e.g. OWL).

The design of the proposed ontology was guided by pinpointing a set of requirements to fulfil by a multimedia ontology that is able to facilitate music producers with semantic metadata to perform the task of annotation of digital music of their choice. The ontology was designed to incorporate support for the music producers so that they do not need to understand the detail of how the acoustic representation of musical audio is associated with musical concepts as understood by music consumers. To achieve that, datatype and object properties were defined to link semantic concepts defined in the ontology. The ontology itself cannot be used directly by music producers to annotate musical items unless it is supported with an appropriate annotation tool. For that purpose existing annotation tools were surveyed so that at least one of those tools could be used to enable music producers in the task of annotation. As detailed in the literature survey, none of those tools were adequate to utilize the semantic knowledge presented by the proposed ***mpeg-7Music*** ontology and none of the existing tools were able to provide adequate support for music producers' annotation. So, a customized annotation tool was designed to test its usability as mentioned in section 3.4.7.

To summarize the plan for evaluating the ***mpeg7Music*** ontology, I have defined a multidimensional perspective that will be presented in detail in section 4.2. It starts by validating the requirements of a multimedia ontology in case of ***mpeg-7Music*** ontology. Next, we will check its structural consistency as ontology. Finally we will present an evaluation of the Simple Semantic Annotator tool as well as the scope of ***mpeg-7Music*** ontology from usability perspective.

4.2. Evaluating the Ontology

Ontologies are usually evaluated from three perspectives - functional, structural and usability profiles (Gangemi et al., 2006). Ontologies are semiotic objects that may be structurally evaluated by looking at it as an information object by checking the formal semantics of the ontology topologically and as well as consistency of the semantics represented by the logical properties. Functional dimension focuses on the intended conceptualization specified by ontology and specifically looks at it as a language as well as its components. From the usability viewpoint, it is important to look at the ontology profile (annotations) that typically addresses the communication context i.e. pragmatics of an ontology. Due to the complexity of content and structure that characterizes music objects, ontology driven representation of semantics concerning music audio prompted us to evaluate the proposed music annotation ontology from three different perspectives. Firstly, we will consider the functional perspective as the research activities, approaches and applications concerning multimedia ontology development have increased rapidly we will consider evaluating the developed ontology with respect to the common multimedia ontology framework in order to promote sharing and harmonization. Secondly to perform the structural evaluation we will validate the proposed ontology using prominent ontology validation tool to ensure its logical consistency. Finally to evaluate the usability profile, we will show the applicability of the proposed ontology using a custom built annotation tool. In essence we will use the following three dimensions for evaluating the proposed ontology in the next three sections:

1. Standard requirements for multimedia ontology - Functional dimension
2. Validating the ontology using prominent ontology validator- Structural dimension
3. Demonstrating the applicability of the proposed ontology using a custom built Annotation tool - Usability dimension

4.2.1. Standard Requirements of Multimedia Ontology

The design of the ***mpeg-7Music*** annotation ontology was guided by a set of requirements. The set of requirements as specified below are based on the general requirements for Common Multimedia Ontology Framework as proposed by Saathoff et al. (2006). I will now show how those requirements were satisfied in the design of the ***mpeg-7Music*** (which is a multimedia ontology) that is the main contribution of this thesis. For a multimedia ontology to be interoperable and to provide prospect for standardization it must conform to the requirements for a common multimedia ontology framework (Eleftherohorinou et al., 2006).

1. **Representation of content structure and reasoning support:** Audio segments can only be decomposed in time or media (MPEG-7 Part 5, 2003). Audio resources should be defined to be part of several audio segments or whole of audio with respect to time considering the media entity with reference to media locator and media duration. So, I created the timeline concept and its subclasses to handle this issue.

2. **Representation of low level features/ roles:** Low level audio features should be connected with higher level concepts to form the definition of harmonicity and percussiveness of sound timbre by definition of appropriate roles.
3. **Support for easy retrieval of metadata:** The idea of metadata retrieval was based on the fact that low level audio features are easy to extract from media files on the fly with the use of MPEG-7 encoder tools (e.g. MPEG-7 Audio Encoder, (2008); Crysandt (2005)) that can create XML metadata. I used this XML metadata as a resource to associate with semantic concepts present in the ***mpeg-7Music*** annotation ontology.
4. **Simplicity:** The music annotation ontology must be a simple ontology that can be used independently to annotate and the generated annotations should facilitate in effective retrieval digital music with reference to the designed ontology.
5. **Separation of Information object:** The design of the proposed ontology was based on ABC MPEG-7 data model with a view to distinguish information about annotated information object from information about content of annotated information object. (e.g. Audio fundamental frequency extracted from media file vs. subjective property of audio as perceived by music lovers e.g. melodic motion and shape).
6. **Link to upper ontology:** The ***mpeg-7Music*** ontology extended the ABC ontology features by utilizing MPEG7-MDS upper ontology concepts to ensure the linkage of the developed ontology for further interoperability.

The design of the ***mpeg-7Music*** ontology was guided by above mentioned six requirements. I will now attempt to check carefully the proposed ontology against each of those of mentioned requirements.

The primary objective of the proposed ontology was to enable the user with the ability to annotate music audio files with ontology driven analysis of the music audio and provide the ontology that was aimed to provide effective retrieval with reasoning support. As a result the annotation ontology satisfies scope and usage requirements as mentioned by requirements 1 and 2.

The proposed ***mpeg-7Music*** ontology contained support for multiple formats of music audio file under audio objects class e.g. wav, mp3 etc. providing further scope of adding other available format in future. Thus it described and represented knowledge for diverse information objects as stipulated by the requirement 5. The ***mpeg-7Music*** ontology was created using a bottom-up approach that was derived from MPEG-7 XML description and it was as an application/task ontology rather than domain ontology (Breitman and Prado 2003). Application ontology is different from domain ontology construction themes that reflect precision and formality of knowledge of the given area. But the objective of the ***mpeg-7Music*** ontology was created in this thesis was aimed to enable users (i.e. music producers) to interact with music annotation system functionality and create usage of the machine process able MPEG-7 data by creating association with it. Thus it satisfies requirement 3.

The ***mpeg-7Music*** ontology served as a backbone for semantic annotation application as we designed in section 3.4.6. As the proposed ontology was designed as 'task ontology' for facilitating music annotation I followed adaptation at the specification level while designing the ontology as well as implementation of the semantic annotation application(Farooq et al. 2007). The primary ontology model specification was completed iteratively satisfying requirements 4 and 6 as follows:

Vocabulary Declaration: Concepts were created to represent multiple audio format (e.g. ResourceFormat class), subclasses under TimeLine concepts (to support for music audio's inherent feature for unfolding in time). This satisfies the content structure requirement.

Categorization of vocabulary: Musical concepts that were captured in the proposed ontology were in line with the timbral and melodic dimensions. Music consumers' vocabularies were categorized under these two musical dimensions.

Proper naming and Identification of Relationships and constraints: Several properties were declared with domain and range relationships defined with constrains and characteristics applied to explicitly mention the permitted values that they may take. This satisfies reasoning support requirement as well.

Proper naming and identification of data characteristics: This satisfies the requirements of representation of MPEG-7 low level features/ roles. This supports the separation of information object requirement as well.

Verification: The iterative specification of the model was verified to check if the model satisfies the simple representation and links to upper ontology requirements.

Support for simple and easy annotation with MPEG-7 metadata: These were checked when the primary ontology model was refined during the design and implementation phase of the semantic annotation tool.

Having traced back to the requirements for a multimedia ontology to be satisfied by ***mpeg-7Music*** ontology I will now focus on its structural soundness in the next section.

4.2.2. Ontology Validation

To validate the structure of ontologies, usually a tool is used; and generally such a tool is known as 'Reasoner'. The reasoning tasks carried out by the reasoner that are considered to be of utmost importance are ontology coherency and classification (Bechhofer 2003). To check **ontology coherency,** the reasoner takes as input the ontology and for each concept in the ontology, depending on whether the corresponding concept is satisfiable or unsatisfiable, it returns true or false respectively. Similarly, for **classification** purpose, the reasoner takes the ontology as input and returns the inferred classification of the various concepts, as opposed to the explicit defined ones. Moreover, for every concept in the inferred hierarchy it also checks if it is satisfiable or not.

Overall, reasoners are used to classify the ontology to show desirable results. Reasoners are pieces of software that implement algorithms to check subsumption, satisfiability and consistency for OWL ontologies. The state of the art reasoners are also known as rules engine or reasoning engine or Description Logic Reasoners and are aimed to check the inference rules commonly specified using ontology language (Fahad et al. 2008) to ensure soundness of the ontologies.

Among the existing semantic reasoning engines Pellet, RACER, and Fact++ are the most popular and effective (Huang, 2008). These reasoners have got notable similarities and differences[20] based on their system architectures, features, and overall performances in real world applications. A further detail can be found in (Delias et al. 2007).

Now, Protege provides open source APIs[21] to use existing reasoners to integrate in semantic web applications. I have used Protege[22] tool (version 3.4) to build the ontology. Besides, other reasoners, Protege provides Pellet for consistency checking of the ontology. Pellet is also a free ware tool to check OWL mark-up for problems beyond simple syntax errors; that will examine the OWL content of the proposed for a variety of potential errors and reports them along with the location of the errors in the files. When using the direct Pellet 1.5.1 reasoner distributed with the Protege 3.4, the ***mpeg-7Music*** ontology appeared consistent and it classified the inferred concepts without any error. The figure 4.1(a) and 4.1(b) shows the result of consistency and classification check report of ***mpeg-7Music*** ontology respectively.

It was mentioned earlier that for the purpose encoding the ***mpeg-7Music*** ontology in OWL and rules specification in SWRL, I used Protege 3.4.1 (version that supports SWRL Tab). After completing the design of the ontology structure and rules specification using Protege 3.4.1 we opened the Pellet plug-in from the Protege Reasoning menu. The reasoning menu opens drop-

20 http://en.wikipedia.ore/wiki/Semantic reasoner#Reasoner comparison
21 http://proteRe.stanford.edu/plugins/owl/api/ReasonerAPIExamples.html
22 http://protege.stanford.edu/

down list for both consistency and classification checking for the ***mpeg-7Music*** ontology file (***mpeg-7Music.owl***). By choosing consistency check option the Pellet reasoner log was generated as shown by the screen shot in figure 4.1(a). Similarly, another screen shot was generated for classification check as shown in figure 4.1(b).

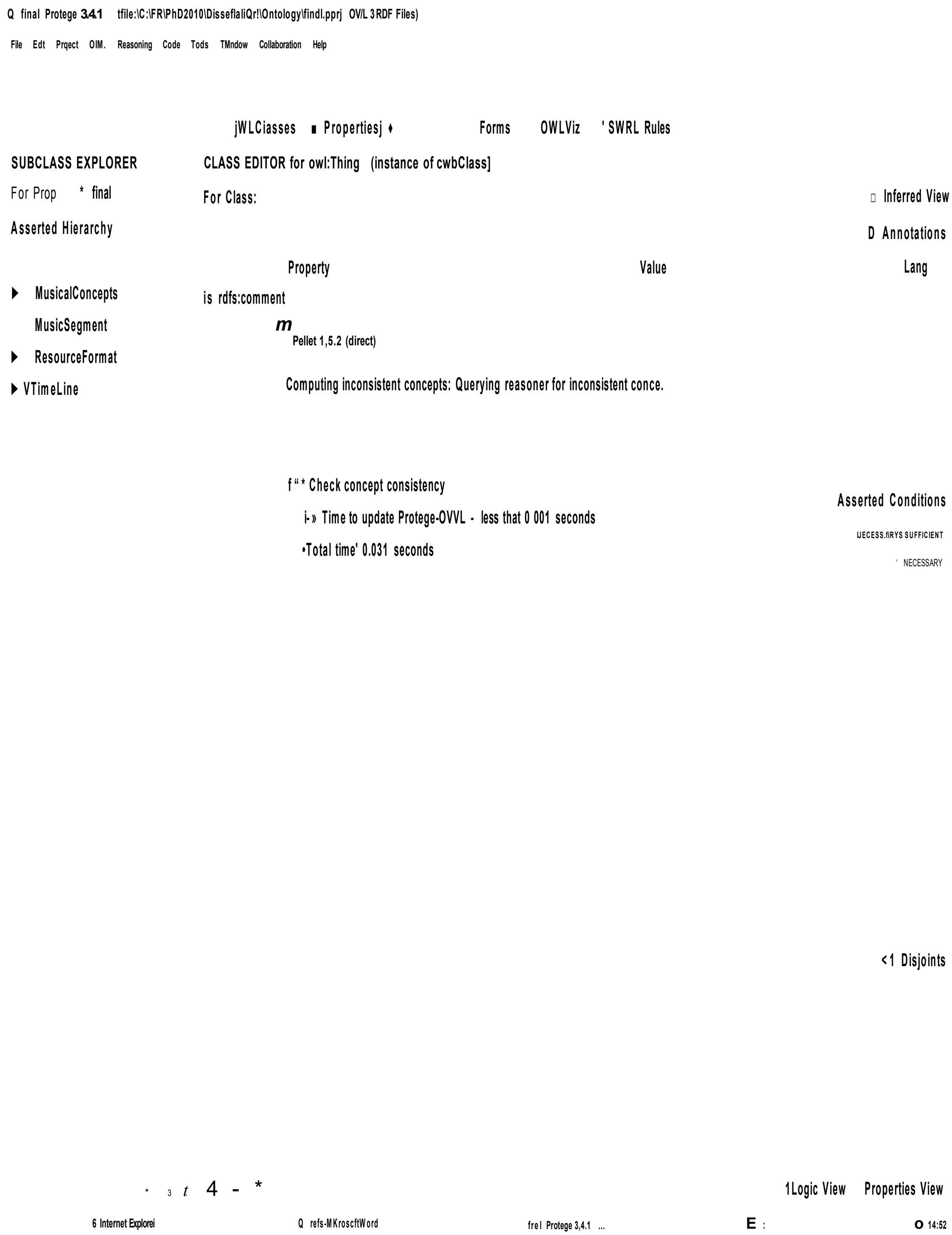


Figure 4.1 (a): Consistency checking by Pellet 1.5.2

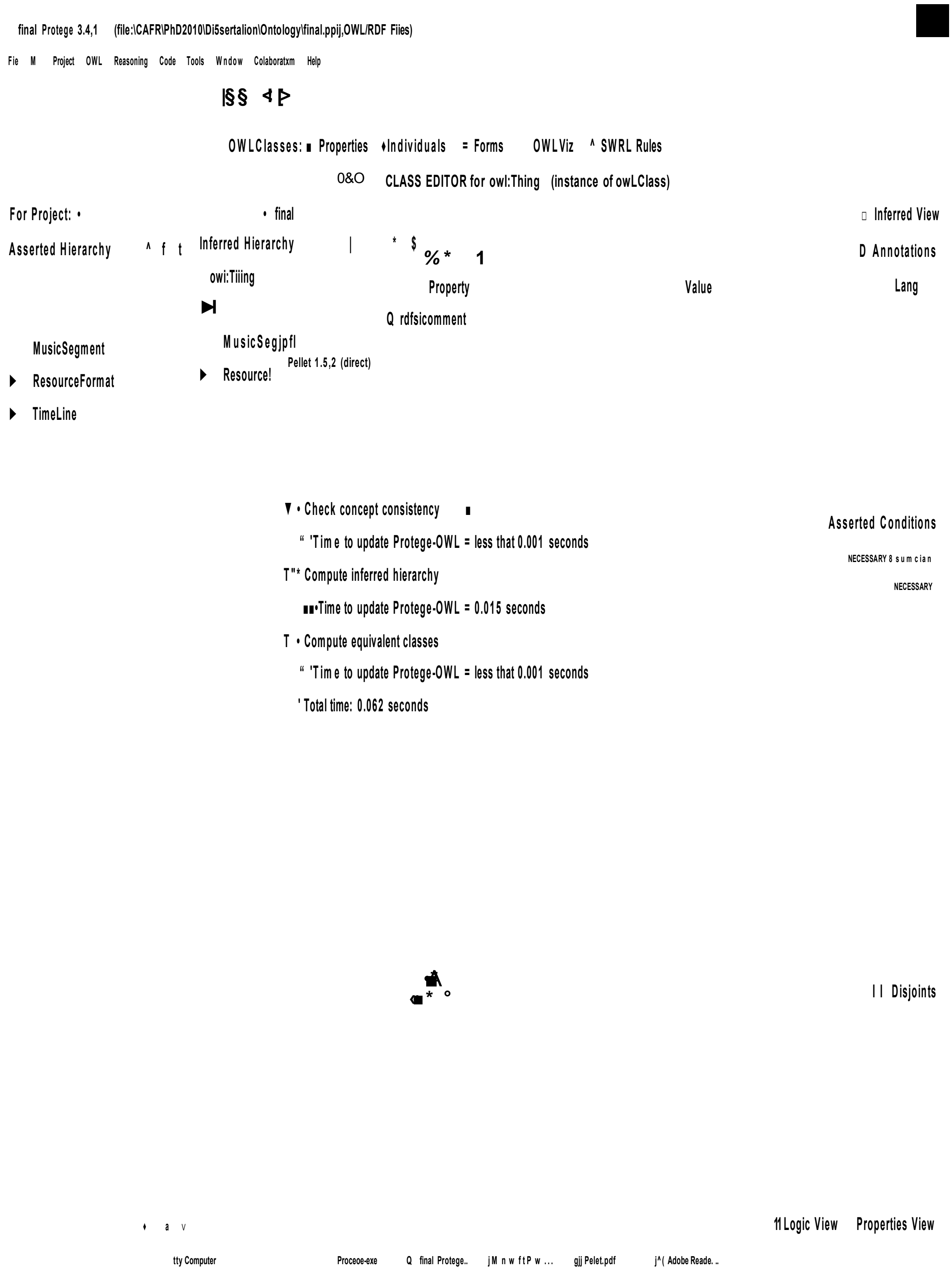


Figure 4.1 (b): Classification checking by Pellet 1.5.2

4.2.3. Demonstration of Usability

Surveying the existing annotation tools available for multimedia annotation a customized tool was developed to evaluate the applicability of the proposed annotation ontology. It provides the music producers with the ability to annotate by connecting music audio's acoustic features with ontological semantics; these auto-generated annotations form the knowledgebase for search algorithms to generate recommendations/suggestions and for music producers to proceed with the task of annotation. The design of the simple semantic annotation tool was focused on three main requirements to be fulfilled by such a tool (section 3.4.6). First, the tool should provide a platform to support annotation of heterogeneous formats of digital music; secondly, the system should provide a standard way to store the created annotations and thirdly it should create an ontology based annotation process. I will now evaluate if the mpeg7Music ontology has proved its potential to be used in simple annotation platform to empower music producers to annotate digital music.

Supporting heterogeneous format of digital music: The ***mpeg-7Music*** ontology formalizes the content format of different category of audio resources and it was designed to guide the annotation process. This ontology works as a knowledge unit that we want to associate with the resource. The ResourceFormat concept in the class hierarchy contains subclasses to support multiple audio formats of music files. More subclasses may be added in order to accommodate more resource formats if needed in future. At present, it provides support for three formats of music files namely, mp3, windows av format and real audio (ra).

Actual Annotation Process: The annotation process of a musical object using the simple semantic annotation tool involves both automatic and semi-automatic association phases. Automatic annotations are generated using the MPEG-7 descriptors mapping that was detailed in section 3.3. Figure 3.15 (c) shows an example of the auto-generated annotations marked using maroon rectangle. The semi-automatic annotations were carried out (by music producers) as shown in figures 3.15 (a-e). The user may upload the music file and name it as shown in figure 3.15 (a-b). The step 3 of the annotation process (figure 3.15(c)) displayed the property hierarchy of the ***mpeg-7Music*** from where the user may choose a property by clicking on it and then step 4 of the annotation process appeared. In step 4 (figure 3.15(c)), the class hierarchy of ***mpeg-7Music*** ontology was displayed. As user clicked on the class concept of choice then individuals of the corresponding class was displayed and the user could choose from the list of those individuals. Once the 'Done' button was clicked then the uploaded reseource was linked by the chosen property (in step 3 of annotation) with the individual phrase (chosen in step 4) structured under ***mpeg7Music*** ontology. So, this semi-automatic annotation process was based on two fundamental annotation steps - i.e. Step 3 that related the chosen property from the ***mpeg-7Music*** ontology with the uploaded music segment; Step 4 in which appropriate class's individuals were discovered and then the user could assign the chosen individual to the music segment. These two fundamental steps were supported by display of and interaction with class & property hierarchy as well as individuals - such functionalities were enabled by the web service client interface used by the semantic annotator application. The web service client interface was served by the web service that was developed for processing OWL ontology.

Storing Annotations in a standard way: According to Li et al. (2007) semantic annotations facilitates semantic retrieval process and ontology based retrieval system requires storing the annotated data with the semantic tag according to the ontology model. So, I designed to store created annotations in OWL format with respect to the ***mpeg7Music*** ontology as entries to MySQL database table.

Usability Testing: As discussed in section 3.4.6 about Web2.0 applications, the popularity of social tagging was a result of the low effort required for tagging. This high usability encountered in social tagging served as the blueprint for the Simple Semantic Annotator. The Simple Semantic Annotator was designed to enable the music producers to generate annotations, and attach them to a music resource. These annotations were metadata containing references to the ***mpeg-7Music*** ontology. The critical success factor in the development of the Simple Semantic Annotator is the usability as detailed in the section 3.4.6. It was designed following the usability standards in the ***Web 2.0*** set for the so-called ***social***

tagging (Bertin-Mahiex et al., 2008). The high level of usability is considered to be a pre-condition for recruiting the end-users' (Music Producers) effort to annotate resources for the use of others (Music Consumers).

Table 7: Evaluation criteria for the Simple Semantic Annotator Tool

Validation criteria	Measurement method	Expected results
Easiness of use (usability)	Average time spent on annotating a resource	Annotating time measured in seconds
	User feedback	Level of satisfaction with the usability

The evaluation of the Simple Semantic Annotator tool was carried out based on criteria mentioned in the Table 7 involving participants who were graduate students not necessarily coming from computer science background. The evaluation session started with six participants to assess the usability of the annotation tool based on the criteria mentioned in table 6. The 5 out of 6 participants took less than a minute to finish the five steps of annotator application to annotate a single music item with single object property. Highly satisfactory feedbacks were received from the participants with regard to easiness of use of the annotator application. The objective of the evaluation session was not only to validate the annotation tool but also to collect recommendation for future improvements that have been discussed later in section 5.3.

4.3. Summary

This chapter presented detailed evaluation of the proposed ***mpeg-7Music*** ontology as well as usability testing of the supporting Simple Semantic Annotation Tool. The ontology was validated against the requirements set out in accordance with the requirements postulated by the Common Ontology Framework proposal (Eleftherohorinou et al., 2006). Then the structure of the developed ontology was validated for consistency and classification satisfiability.

The proposed Simple Semantic Annotation Tool was built in order to show the scope of the mpeg-7Music ontology in guiding the annotation task. The tool was developed based on Web2.0 validation criteria and the evaluation of the tool by users provides high degree of satisfaction in terms of usability.

5- Discussion and Conclusion

Annotating music items poses specific challenges within the field of semantic tagging/annotation of multimedia. This is due to many possible descriptions and interpretations of one musical item by music consumers and to the semantic gap between what can be automatically derived from the raw data of a musical audio resource. The focus of this thesis was on the process of annotating musical items that is based on music consumers' information needs and annotation. Problems related to annotating musical audio resources were studied and solutions were developed that were based on extending the semantics of musical audio resource descriptions through background knowledge present in the ***mpeg-7Music*** annotation ontology and audio-analysis techniques of MPEG-7 feature extraction tools.

So, the main contribution of this thesis is the ***mpeg-7Music*** annotation ontology and its main components are:

- A structured vocabulary for annotating digital music i.e. ***mpeg-7Music*** ontology with a set of class concepts and property roles
- A framework for annotating digital music- a customized annotation tool
- Methods to combine structured representation of music consumers' unstructured tags with MPEG-7 audio analysis

The next two sections revisit the research question in order to find out how much this contribution was able to address that and then a critical evaluation of the outcome of this thesis is presented by assessing its impact in the related area of research. The final section presents conclusion with future research directions.

5.1. Contributions in terms of the Research Objective

In this section, I have attempted to revisit the goals and research questions raised (How to annotate music against MPEG-7 description to deliver meaningful search results?). It starts by discussing methods used and results achieved.

The basic question addressed in this thesis is how to annotate music against MPEG-7 description to deliver meaningful search results. This question prompts to handle two distinct aspects of the problem of annotation of music. First one is how to annotate music using MPEG-7 description and the second one is how to represent the MPEG-7 enriched annotations for delivering meaningful search results.

Generally, tagging/annotating of music are done by diverse types of users and the created annotations are very versatile in nature as these are dependent on who is annotating the music and what are the purposes for doing it. Web 2.0 folksonomies are created by end-users (both consumers and producers) tagging could be a cheap and quick remedy to address the first part of the question if end-users would be able/interested to utilize the MPEG-7 features for tagging music of their choice. But, except the acoustic experts it is quite impossible for end-users to utilize these features. Other than engaging music consumers with the task of annotation, this thesis considered targeting music producers to deal with the task of annotation.

MPEG-7 description of music audio content can be generated automatic feature extraction tools; so automatic annotation tools could be a straight forward solution to generate annotations with MPEG-7 features but it requires huge training of the machine learning algorithms and the generated annotations would lack the high level semantics that may be created and understood by end-users.

Now, looking at the second part of the research question that suggests to enrich MPEG-7 with semantics in an attempt to provide meaningful search results, Semantic Web technologies seems to be an ideal candidate to model MPEG-7 features. But the big issue to resolve is how to enable music producers to annotate and create a knowledgebase that can be used by search algorithms. Automatic annotation tools suffer from the old problems related to the so called *semantic bottleneck*: development and maintenance of high quality real life knowledgebases is costly, tedious and error prone. Web 2.0 with its folksonomies did not provide techniques to develop meaningful annotations and did not enable search engines to use them and manipulation of folksonomies by machines appears down to the complexity of the natural language processing. Folksonomies can be utilized as a means for cheap and quick alternative for automatic annotation but they are not designed for machine processing.

So, I was left with the challenge of how the insights from Web2.0 examples could be used to develop a semantic annotation tool and enrich MPEG-7 with meaning by the music producers. To deal with this challenge, a set of unstructured music consumers' tags were chosen to model the semantic vocabulary using semantic web structured vocabulary encoding language (i.e. OWL 1.0) to formulate a machine process-able representation of those tags and by defining concepts and properties to link MPEG-7 data type features and create combined and structured meaningful vocabulary. Looking at the pros and cons of both fully manual and automatic annotation tools, a semi-automatic annotation tool was designed to empower music producers to annotate music items without the need for understanding the acoustic metadata represented by MPEG-7 feature description.

So, the research objective was achieved by enabling annotation with MPEG-7 features that are mapped with the proposed semantic metadata vocabulary i.e. ***mpeg-7Music*** ontology. It fulfils the goal of lifting MPEG-7 metadata to semantic metadata structure. This ontology provides support to generate meaningful (machine process-able) annotations by means of its OWL encoded metadata structure. Utilizing the ***mpeg-7Music*** annotation ontology with the customized semi-automatic annotation tool for music producers creates a trade-off to bring the best of both worlds of Web2.0 and Semantic technologies.

5.2. Impact of the proposed music annotation ontology

The literature survey (Chapter 2) presented the state of the art in, hereafter termed as, research inquiry areas; -

a) search and retrieval of music,

b) semantic annotation of music and

c) multimedia description standards.

My observation of the existing solutions showed that the efficiency of music search and retrieval is largely dependent on music tagging efforts. But, mere tagging produces collections of keywords and those keyword collections are generally unstructured. Search techniques that work on the unstructured free-form keywords that are called syntactic metadata are only best enough to provide syntactic search as described in section 1.1. This thesis envisioned and planned to generate a set of structured metadata from music consumers' tags by taking inspiration from the semantic web technology's idea of producing machine process able metadata or in other words meaningful (semantic) metadata. Then the techniques to represent semantic metadata were investigated and found that the Web Ontology Language (OWL) as the most dominant language to model the semantic metadata. Before further exploration on the detail representational issues, I started searching/examining annotation tools to demonstrate the applicability of the envisioned semantic metadata for music in a way that actually would enable music producers to annotate/tag music items. The search for annotation tools was guided by three minimum criteria - **i)** support for annotation with reference to OWL ontology, **ii)** standard representation of annotations and **iii)** ability to enable annotation for heterogeneous digital format of musical content. Having consulted contemporary survey literature on annotation frameworks and independent evaluation of existing tools, it was observed that no single tool that could be borrowed to demonstrate the scope of the semantic metadata vocabulary i.e. ***mpeg-7Music,*** the proposed OWL ontology for music annotation. Inspired by the success of Web 2.0 applications that produced tagging functionality to engage ordinary users to tag music of their choice, the simple annotation tool was designed so that it satisfies three aforementioned criteria (i-iii).

Again, looking back to representational standards of music metadata, existing initiatives towards music ontology, multimedia ontology and MPEG-7 audio features (not in any order) were explored. MPEG-7 Audio standard provides a set of metadata that can be automatically generated from underlying music audio. Related research on metadata generation from MPEG-7 features did actually follow two separate directions as depicted in figure 5.1.

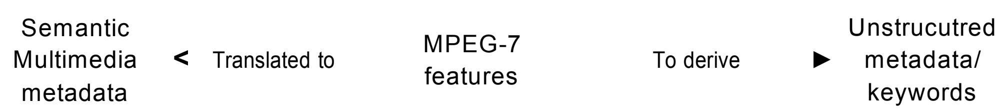


Figure 5.1: Metadata Generation from MPEG-7 features

In connection to MPEG-7 audio features of musical objects one thread (marked by green block in figure 5.1) of research investigated how MPEG-7 audio features may be interpreted directly to classify music with keywords like sweet, nice etc. (Whitman, 2004) in an attempt to derive unstructured/syntactic metadata from standard MPEG-7 audio features. In the context of the problem of semantic search of music this could not provide a solution to eliminate limitation of keywords based searching. There was another attempt that applied a mapping of their customized keywords with MPEG-7 audio features to derive a set of keywords to describe music belonging to a particular culture (Tsekeridou et al., 2006). This also belongs to the same thread as it derives unstructured metadata from MPEG-7 features. The other thread (marked by magenta block in figure 5.1) started creating a parallel semantic vocabulary to translate MPEG-7 constructs describing multimedia. This only gives us a general set of semantic vocabulary to represent MPEG-7 metadata in a meaningful way understood both by humans and by machines. Attempts under this thread did not provide solutions about how to create a link of the semantic metadata generated by them with corresponding MPEG-7 features and the most important missing part in the context of this thesis was how these semantic multimedia metadata could be effectively utilized to put in the use of semantic search and/or semantic annotation of music. Following this same thread, it was found that existing efforts to create visual ontology to represent image and video but none was specifically designed for music representation in this context of semantic multimedia metadata.

Now, as far as the music ontology is concerned it was identified that the initiative to design music ontology was rooted from the idea of providing a structured ontology format for the metadata needed to describe musical performance from editorial point of view and record keeping purposes. It had nothing to do with digital audio content representation standards like MPEG-7 that may be used for automatic search and retrieval of music.

To justify the impact of the music annotation ontology in three different research area (a-c) as mentioned in the beginning of this section –a critical discussion will be presented next about the main contribution which is an ontology that is designed to facilitate as a backbone to perform the task of semantic annotation by music producers; by identifying the similarity and contrast of ***mpeg-7Music*** ontology with existing efforts by evaluating its impact in those research areas. In this connection, the issue of automating semantic annotation process as an innovation using Web2.0 technology will also be discussed.

To summarize the plan for critical evaluation of this contribution I have chosen a five point discussion and each one or more of these points has got its root with three (a-c) broad research fields have been outlined before. For example, unstructured music tags are not able to provide efficient search and retrieval of music- causing limitation to the field of semantic search; current music annotation tools cannot utilize different levels of information present in musical items - showing a clear shortcoming of the semantic annotation research theme; besides, standard multimedia semantics do not contain sufficient structured concepts and properties to take advantage of different levels of information present in music objects (specifically acoustic information) - draws to the need for more study with standard multimedia semantics. Now, such points will also be presented to qualitatively evaluate the proposed ***mpeg-7Music*** annotation ontology as detailed in the next five subsections.

5.2.1. Structured vocabulary for music tagging

The current state of music tagging fails to utilize the different levels of content information that is encapsulated in the musical object - as the tags created by the users are only applicable at the knowledge level (highest level) as showed in the figure 3.1. This actually leads to ambiguity for the search algorithms when such tagged information is used for enhancing search results. Thus, existing music tagging approaches suffers from the problem of low precision and recall as the created metadata is aimed to be analyzed completely at the consumption level as depicted in figure 2.3.1believe that such problem could be minimized to some extent if metadata would be categorized before it is made available for search engine consumption. Because, rather than a free form vocabulary, a structured organization of metadata vocabulary could provide us with effective search results. Besides, structured semantic vocabulary can leverage the use of multi-level knowledge presented in the music items (e.g. signal level data wrto. figure 3.1). Social tags' folksonomy (Weller, 2007) based representation of music tags contain weakly labelled unstructured free for all vocabulary. Work presented in (Whitman, 2004) concentrated towards deriving unstructured keywords from MPEG-7 feature description of musical audio. Similarly, mapping of MPEG-7 metadata description of musical audio to derive culture specific music metadata (Tsekeridou et al., 2006) led to the generation of unstructured syntactic music metadata.

Innovations related to music metadata generation using only Web 2.0 technology and utilizing the prospect of creating standardized set of metadata with MPEG-7 are yet to bring us to the state of having a structured semantic vocabulary for music. Besides, outcomes of existing attempts towards creating music metadata are basically producing syntactic metadata that are not meaningful in the context of semantic web's vision of machine process-ability (Gruber, 2008). The proposed ***mpeg-7Music*** annotation ontology will overcome the limitations of current state (Turnbull et al., 2008) of music tagging (section 2.1.3) that fail to utilize the different levels of content information that is encapsulated in the musical object. At present there is no structured vocabulary that lifts low level acoustic property of music audio needed for effective tagging. This ontology will enhance current approaches of music tagging by creating a knowledge based representation of musical content that will provide implicit association with low level acoustic properties (in MPEG-7 audio format) of musical sound with music consumers' keywords so that the music producers will be able to tag the musical object without being aware of complex scientific representation of low level features. Besides, ***mpeg-7Music*** ontology was encoded using semantic web's OWL standard. The designed annotation tool provides a framework for annotating music with low level features that are easy to extract using automated tools.

5.2.2. Semantic Search and Retrieval of Music

At present, search engines consider semantic search to be implemented during analyzing the textual query only, based on the assumption that users are unable to type much more than a simple keyword for searching music as discussed in section 2.1.2. This in fact leads to the challenge of how to represent and index the music item for efficient retrieval.

Music information retrieval (MIR) systems (Baumann et al., 2002) do rely on end-users textual query and structured vocabulary to satisfy search results. Effectiveness of such systems depends on the design of structured vocabulary. One way of providing structured vocabulary for processing search queries is to generate metadata automatically from music resources (Knees et al., 2007) from web pages containing contextual information about music files by combining natural language processing with semantic information related to audio as well as contextual metadata of audio links with low level acoustical features. The other available option (Kim et al., 2004) is to rely solely on generating automatic metadata index from acoustical data. But, automatic metadata generation techniques require training data and often produce unsatisfactory tags. Moreover, on the fly metadata generation and analysis creates a significant efficiency constraint for search algorithms.

The effective option could be providing a structured semantic vocabulary for music search and retrieval. The ***mpeg-7Music*** ontology was designed to annotate music items and it stands as a structured semantic metadata. Music items annotated with reference to this ontology may be made available to MIR systems that will use ***mpeg-7Music*** ontology as its structured vocabulary (as shown in figure 5.2), it represents a MIR system that depends on textual queries to be mapped to formal semantics that will create semantic indices to provide semantic search results. At present, no formal semantics is available to provide required formalization to map users' (textual) query that are intended to search for music items.

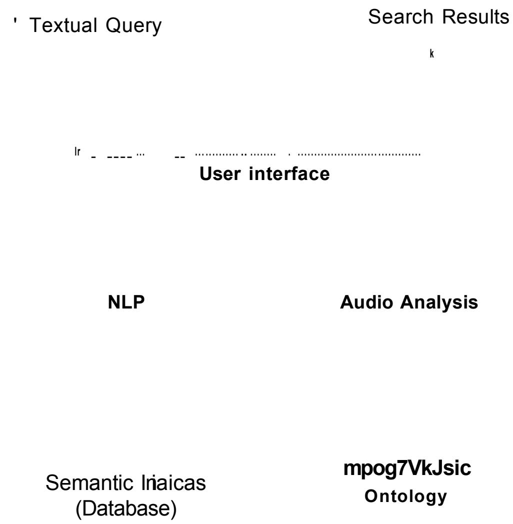


Figure 5.2: Semantic Music Information Retrieval Application

The proposed ***mpeg-7Music*** ontology contains the logical structure that is able to map music consumers' keywords (contained in textual form of search query) to its formal ontological semantics. As a result, it will contribute to provide semantic search and efficient retrieval. The annotations created by the music producers with reference to ***mpeg-7Music*** ontology will be stored as database indices. Textual queries will be analyzed using those semantic indices (those are actually music producers' annotations) with respect to the ***mpeg-7Music*** ontology.

5.2.3. Enriching multimedia semantics with end-user keywords

At present multimedia ontologies only serve as upper level concept ontology (SNOEK et al., 2006; Ceccaroni, 2001; Naphade et al. 2006). Some of the existing multimedia ontologies are confined to a particular domain (Snoek et al., 2006; Luo and Fan, 2006; Bertini et al., 2007; Bertini et al. 2005) but not to the music domain directly. Among the existing ontologies that covers music domain (e.g. Raimond, 2007) fall mainly on the category of task/method ontology and serve for commercial exploitation (sharing, production and recommendation) rather than providing music consumers with satisfactory search results.

There are several MPEG-7 multimedia ontologies that also were designed to provide upper level concept ontologies and were developed to represent the semantics of MPEG-7 feature description. For example, the ABC ontology provides generic terms like AudioSegment (Hunter and Little, 2007) but does not cover the detail semantics that could be represented for music segments of a musical object to capture meaningful representation understood by end-users. The DS-MIRF ontology (Tsinaraki, 2007) describes relations (relation types) that may be further extended for specialized applications but are not sufficient for tasks like semantic annotation by music producers that in turn may be used for efficient retrieval by consumers. Semantically annotating music items is cumbersome for music producers because they are quite unable to interpret high level meaning of their phrases with signal level metadata but these signal level metadata can be automatically extracted from music audio using MPEG-7 feature extraction tools.

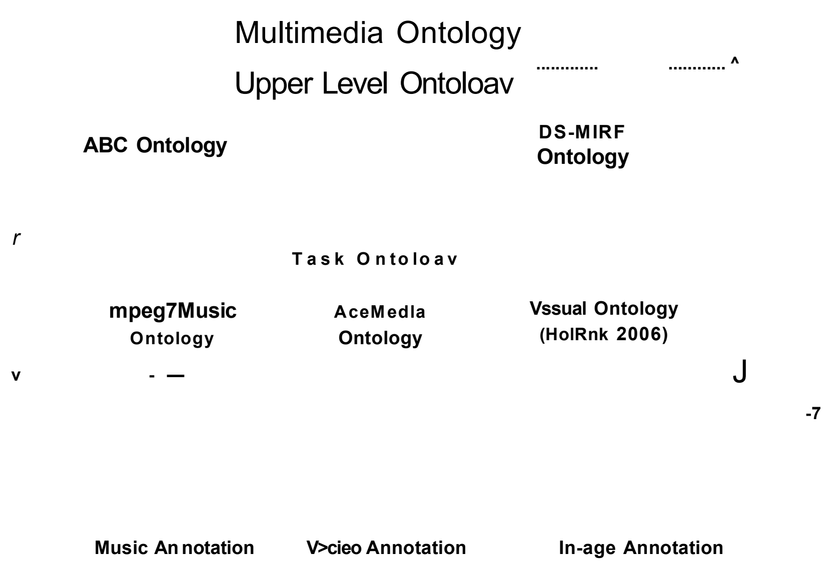


Figure 5.3: Multimedia Ontology types

Again, complete automation of semantic annotation without human intervention yields unsatisfactory outcome. Existing MPEG-7 compliant, upper level multimedia ontology only provides generic terms and properties to describe multimedia in general. These do not support to represent any specific concept to facilitate music producers for annotating music.

Upper level ontologies (e.g. ABC or DS-MIRF) are further extended or re-used by task ontologies (figure 5.3) for different purposes. For example, AceMedia ontology (Petridis et al., 2006a) was designed to provide support for annotating video items. Visual ontology (Hollink, 2006) was specifically designed to annotate images. But, in the contemporary works, I could not find any existing ontology that was designed to capture and represent different level of information from music items and sufficient structured vocabulary for music producers to annotate those. The ***mpeg-7Music*** annotation ontology extends existing multimedia ontologies in two ways: Firstly, it extends the upper level MPEG-7 compliant ABC ontology (Hunter and Little, 2007) by extending the generic Multimedia ontlogy's concept AudioSegment and creates a MusicSegment class and establishes association of music segments with music consumers' keywords by defining MusicalConcepts class hierarchy. This work is closest to the (Hollink, 2006) that creates a visual ontology for image annotation by extending concepts from ABC ontology. Figure 5.3 shows the domain of MPEG-7 compliant multimedia ontologies; where the ***mpeg-7Music*** stays as a music annotation task ontology having its foundation on ABC ontology. Secondly, it considers the existing Music Ontology (Raimond et al., 2007) to incorporate TimeLine concepts with music consumers' tags or phrases used by them. Thus it re-uses concepts from the existing structured editorial metadata collected and organized from folksonomies (Turnbull et al., 2008; Weller, 2007) as folksonomies have been able to attract attention of music lovers and are successful on enabling and engaging ordinary users to tagging music items. Proposed ***mpeg-7Music*** ontology has brought a new dimension to Music Ontology by adding automatically extracted acoustic features (MPEG-7 semantics) to enrich it.

5.2.4. Standard Interoperable representation of Music

Current trend in music tagging follows the use of unstructured keywords to annotate music leading to ambiguity. Because, there is no standardized metadata scheme to tag music and the latest widely acceptable multimedia description standard (MPEG-7) needs to be adapted for music metadata representation usable by music producers. Besides, semantic search of music will require representing the music description metadata in machine interpretable encoding in dominant semantic web ontology standard i.e. OWL.

Current multimedia ontologies seem to be following a quite different direction from the efforts that attempted to derive unstructured metadata and keywords form MPEG-7 features as showed in figure 5.1. These two threads of research took different approaches to utilize the prospect of MPEG-7 initiative. Now, if I redraw the figure 5.1 in an attempt to relate these two threads of research efforts it arrives at the depiction as in figure 5.4 where the dark orange block marks the place of this contribution in joining these two disparate fields of research using the orange arrowed lines.

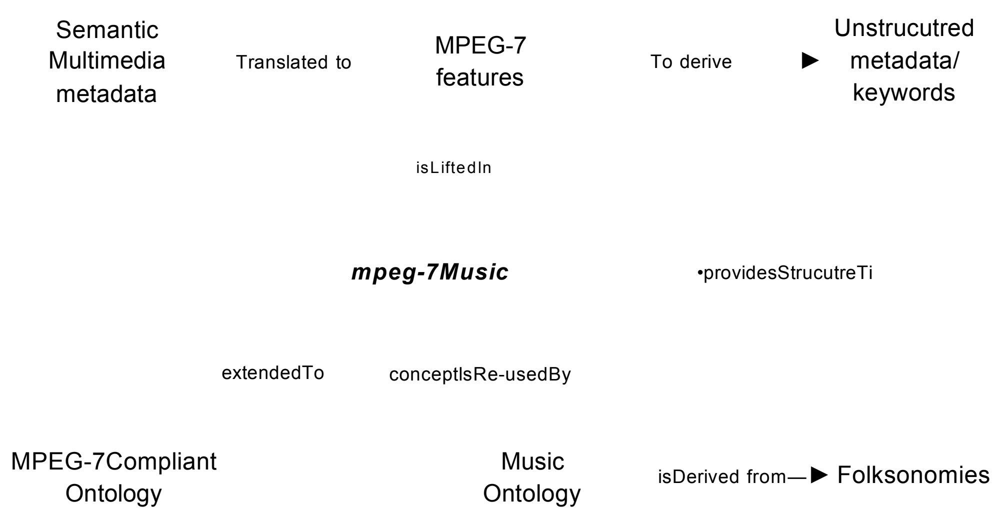


Figure 5.4: Presence of mpeg-7Music Ontology in the research field

Referring to figure 5.4, the MPEG-7 compliant ontologies were a specialized set of semantic multimedia ontology that was designed to represent the semantics of MPEG-7 metadata. The Music Ontology was derived from Folksonomies that comprised of mainly unstructured tags/keywords. The mpeg-7Music ontology utilizes the outcome of these separate research efforts those were not interoperable. It establishes a four way connections among them. It extends MPEG-7 compliant ABC ontology and re-uses concepts from Music

Ontology. It distinctly lifts MPEG-7 features in it to take advantage of auto generated MPEG-7 acoustic metadata. Further, it provides a semantic structure to ordinary music consumers' unstructured keywords by associating these keywords in its ontology organization. To do this, I used a subset of keywords from work carried out mainly by Sarkar et al. (2007) as well as Schmidt-Jones (2010). The proposed ontology has been encoded using OWL 1.0 standard to make it semantically interoperable. Before the mpeg-7Music ontology there was no music representation ontology that supports for music producers' annotation conforming to both MPEG-7 and OWL standards.

5.2.5. Automating the process of Semantic annotation of Music

In the existing literature it was found that automation of semantic annotation is still an open issue (Tsinaraki, 2007). The manual annotation of every single music/ music segment is cumbersome while full automatic annotation is not able to capture semantic information depicted by that music segment.

Annotations are needed to generate metadata to achieve easy and effective search & retrieval of resources/ digital items. But, annotations created manually by attaching unstructured metadata with digital objects lead to ambiguous search results. On the other hand annotations generated by fully automated tools lack in high level meaningful interpretation. So, there are clearly two issues to address - structured metadata for annotation and automating the process of annotation.

In the last two sections, it has been shown that ***mpeg-7Music*** supports for generating meaningful annotation of music items as this has been designed to be a structured metadata in the form of ontology from MPEG7 standardized metadata and the ontology has been encoded using dominant OWL standard.

The next problem to be addressed is how to use this ontology for annotating digital music. Web2.0 technologies (Zhang, 2007) to develop folksonomies did not provide techniques to develop meaningful annotations but could be utilized to automate the annotation process for music producers quickly (Bateman et al., 2006). Annotations generated using Web2.0 applications were not aimed for machine processing and on the other hand, the field of Semantic Web technology is still suffering from semantic knowledge acquisition bottleneck (Gruber, 2008). Consequently, it was decided to use semi-automatic annotation tool for end-users creates a trade-off to bring the best of both worlds of Web2.0 and Semantic technologies (Damme et al. 2007). Now, the rationale behind using a semi-automatic tool came from the observation of consulting and evaluating existing annotation tools that complete automation of annotation process leads to a compromise with the quality of annotated information on one side because such tools are unable to capture the high-level meaning conceived by human users. Fully manual annotation in case of this contribution would not be able to attract/engage users to achieve research objectives of this thesis because of error-prone annotations and the laborious effort needed.

Currently available annotation tools (Uren et al., 2006) fall into two broad classes -i.e. simple syntactic annotation tools and ontology guided annotation tools. I tried to find a tool that would support for semi-automatic annotation by end-users and satisfy three basic criteria. Firstly, the most important criterion is to be the ontology guided annotation because I had to demonstrate the effectiveness of ***mpeg-7Music*** ontology. Secondly, for the annotated information to achieve interoperability through heterogeneous systems the representation of annotations using a standard format is required. So, the ***mpeg-7Music*** annotation ontology was encoded using dominant OWL 1.0 standard and for the same reason it was preferred to preserve the annotated information accordingly. The third criterion considered was the ability of the tools to support for diverse content format of music audio files. My observation shows that a good number of tools fall under the category of simple syntactic annotation tool and ontology guided semantic annotation tool as shown in table 1(a) and table 1(b) respectively. Most of them support annotation of textual content only; very few of them could enable annotation for other types such as MPEG, JPEG, QuickTime, JMF etc. None of them did consider the huge variety of digital music media of different types such as audio (wav, mp3, ra etc.). Among the tools that were categorized as ontology guided semantic annotation tools did not consider support for diverse audio format and standardized OWL encoded ontology guided annotation - both of these two criteria together.

So, to achieve the research objectives to automate the part of the annotation process that lifts MPEG-7 audio features description, map those with the ontology and presents the analysed information as suggestions to the music producers, I had to develop a semi-automatic ontology guided annotation tool that supports them to annotate music files. It provides a unique platform that utilizes semantic association rules of the signal level metadata description of the music audio with music consumers' tags. The ***mpeg-7Music*** ontology serves as the backbone for automation of the annotation process but requires human intervention (i.e. from music producers) to attach meaningful tags.

5.3. Conclusion and Future Directions

The digital music industry is one of the biggest internet based industries worldwide. As part of this research effort it was observed that the current trend of searching for music by using music consumers' keywords/tags is unable to provide satisfactory search results due to insufficient use of underlying acoustic metadata. Search and retrieval of music may potentially be improved if music metadata is created from semantic information provided by association of music consumers' tags with acoustic metadata because acoustic metadata is easy to extract automatically from digital music items.

This thesis created a novel solution based on semantic technologies, showed the potential of ontologies to serve as a backbone for annotating music items and provided scenarios for the application of semantic technologies in the digital music industry. The main contribution under this thesis is the first prototype of ***mpeg-7Music*** annotation ontology. The novelty of the proposed ontology can be justified from different perspectives: firstly, to the best of my knowledge there is no music annotation task ontology that creates a unique opportunity for music producers to annotate music with its audio properties as represented by the MPEG-7 encoding. Existing multimedia ontologies only deal with image and video annotation and do not consider the unique requirements to be addressed for music annotation. Besides, this ontology satisfies the requirements set forth for multimedia ontology. Secondly, from the insights that arose from surveying existing multimedia annotation tool, none of the existing effort considers designing ontology for supporting the annotation task of digital music coming in different content formats and so a customized semantic annotation was designed to demonstrate the applicability of the ***mpeg-7Music*** ontology. Thirdly, this ontology was designed according to the dominant semantic web ontology standard (OWL 1.0) as well as multimedia description standard (MPEG-7 Audio).

A critical evaluation of the proposed ontology was performed by assessing its impact in five different areas where this contribution fills a clear gap. The current state of music tagging could be improved significantly if a structured metadata scheme is used for tagging by music producers– so, ***mpeg-7Music*** was structured as ontology to provide meaningful metadata. Music information retrieval systems implementing semantic search techniques also require structured metadata to provide satisfactory search results against textual query. The proposed ***mpeg-7Music*** ontology may also be used in the music search engines to fulfil the requirements for structured metadata. Existing MPEG-7 compliant multimedia ontologies are not appropriate for use by music producers' annotation.

The ***mpeg-7Music*** ontology creates a bridge with music consumers' tags and MPEG-7 acoustic metadata and extends upper level multimedia ontology i.e. the ABC ontology and the Music ontology (Raimond et al., 2007). This ***mpeg-7Music*** ontology was designed by creating a mapping with MPEG-7 Audio data types and was encoded in OWL1.0 syntax and thus it creates a standard interoperable representation.

Finally, the ***mpeg7Music*** ontology which is the main contribution supported by a semi-automatic tool have demonstrated their potential to stand as light weight concept ontology for annotating digital music by music producers. This effort provides a novel contribution towards bridging the connection between MPEG-7 compliant ontologies and the Music Ontology establishing a four way connections between these two– first by extending MPEG-7 compliant ABC ontology and re-using concepts from Music Ontology; secondly, it distinctly lifts MPEG-7 features associating semantics with MPEG-7 acoustic metadata; thirdly, it provides a semantic structure to music consumers' unstructured keywords by organizing these keywords into ontology class instances. Moreover, it extends upper level MPEG-7 multimedia ontology creating interoperable music annotation ontology in dominant ontology representation language like OWL.

The proposed ontology as detailed in section 3.4.5 is the first prototype of ***mpeg-7Music*** ontology. This thesis envisions that the proposed ontology can be further improved in future in several ways as mentioned in the following paragraphs. Besides, as a further consequence of the critical discussion of the impact of the research carried out in this thesis, some directions for future R&D are also discussed below.

At present the ontology contains only two association rules to map *timbralDescriptor* and *melodyDescriptor* datatypes to two direct subclasses of MusicalConcepts class only. To this date there is insufficient research evidence to formalize datatype descriptors' exact values to assign with instances of musical tags used by end-users. So, the association rules are left to map to only upper classes (such as brightness and motion classes). As this ontology was scoped to act as a backbone for annotation, the proposed annotation tool will be used to capture more annotated data by disseminating it for music producers to annotate more music items. When sufficient data might be gathered then more rules may be defined using the SWRL tab that have been mentioned before in section 3.4.5.

Due to the imprecise and uncertain nature of the music annotations produced by end-users it was not possible to define precise formalization of rules in the proposed ontology. Only few association rules were declared in this thesis to link MPEG-7 descriptors with music segments using property definitions. To define more rules to create associations like how bright it is and what does "nasal" or "flute-like" quality mean as expressed in numbers by the descriptors it requires to depend on the relationship of the subjective parameters (brightness/sharpness of sound) with objectively derived values from acoustic parameters (Kostek, 2003). Such association varies from user to user. So, rules for explicitly representing such phenomena will require us to define fuzzy rules (Pan et al., 2006) that are similar in form to SWRL rules (antecedent -> consequent), except atoms in both the antecedent and consequent can have weights/ importance factors using numbers between 0 and 1.

To be able to populate the proposed ontology with fuzzy rules, at least two things are needed. Firstly, to acquire weight parameter associated with timbral characteristics (e.g. fuzzification of timbral concepts such as brightness) we have to consider research findings from (Kostek, 2003) or perform enough independent survey with music consumers. Secondly, to model the rules, availability of SWRL-like rule editor (Fudholi et al., 2009) that support for defining fuzzy rules must be considered.

This thesis has modelled only two musical dimensions in this ontology. In future, further work will encompass adding more classes and properties in this ontology to conceptualize other musical dimensions e.g. rhythm. This work aimed at utilizing only those MPEG-7 audio features those are automatically extractable by existing MPEG-7 feature extraction tools. Rhythm modelling requires tempo / beat measure for which it requires to use beat counter tool that could be integrated with the designed semantic annotator tool. So, to incorporate rhythmic concepts in this ontology, an interface needs to be developed for the beat counter data using available beat counter softwares[23] that might be interfaced with the annotation tool to count beat per minute data from the music file; at the same time ***mpeg-7Music*** annotation ontology will also need to be added with concepts and properties to interlink beat per minute information presented by the music file.

The proposed ontology can be evaluated further by implementing a semantic search interface as showed in figure 2.7 to validate and test it in conjunction with the stored annotations. This will require us to design and develop an indexing server to rank the search results with reference to the proposed ontology.

The future extensions that have been identified here provide new project ideas for semantic web application concerning music search & retrieval. The generated annotations by the annotation tool were stored in a MySQL database on which the search and retrieval applications may operate to create the indexing server in order to rank the search results. Then the generated indices may further be utilized with reference to the proposed ontology in this thesis to retrieve the music item of choice. So, the proposed ontology which is the main contribution under this thesis works as the backbone for the designed Semantic Annotator prototype and foundation for future improvements.

The evaluation of the impact of the ***mpeg-7Music*** ontology as summarized in section 4.3 prompts us to list future R&D insight. The section 5.2 has presented a critical discussion of the contribution in the field of semantic music search and retrieval showing a schematic diagram (figure 5.2) of how the proposed ontology can be used as a reference for semantic knowledgebase for music search engines. Implementation of music search engine interface

---

[23] Vocalist.org.uk, 2001-2010. [online] Available at: <http://www.vocalist.org.uk/beats_per_minute_reference_chart.html#free_beats_per_minute_software>[Accessed 17 April 2009]

and efficient search algorithms to rank semantic indices are also included in the future research interest.

The semantic annotator application may also be complemented with audio segmentation functionality as part of its automating the process of annotation of music (in addition to section 3.4.6) by enabling the users to create several segments of a single piece of music and annotate those segment with concepts from the proposed ***mpeg-7Music*** music annotation ontology. Supplementing the semantic annotator application designed here would create further opportunity for us to evaluate the ontology more comprehensively.

Last but not the least; a further plan is to make this ontology available for public use by publishing it for MusicBrainz or LastFM users so that it may be tested with ordinary users (both music consumers and publishers). Publishing this ontology, may require us to customize it before it is made available to be usable in a public domain. At present the ***mpeg-7Music*** ontology contains only a few instances (music consumers' phrases) as presented by Sarkar et al. (2007) and Schmidt-Jones (2010). Depending on the need of the users of MusicBrainz and LastFM it would be required to add more instances or revise the existing instances of the ***mpeg-7Music*** ontology.

References

[Abels et al. 2005] Sven Abels, Liane Haak, Axel Hahn. Identifying Ontology Integration Methods and their applicability in the context of product classification and knowledge integration tasks. Technical Report: Department of Business Information Systems University of Oldenburg. Report No. WI-OL-TR-01-2005. Date of Publication: October 2005.

[Agostini et al. 2003] Giulio Agostini, Maurizio Longari and Emanuele Pollastri. Musical Instrument Timbres Classification with Spectral Features. EURASIP Journal on Applied Signal Processing 2003:1, 5-14

[Ahn and Dabbish 2004] L. von Ahn and L. Dabbish. ***Labeling images with a computer game.*** SIGCHI 2004.

[Ali and Aarabi 2006] Ali, S.; Aarabi, P. A Novel Interface for Audio Search. IEEE International Conference on Multimedia and Expo, 2006.

[Antoniou and Hermelen 2004] Grigoris Antoniou and Frank van Hermelen. A Semantic Web Primer. 2004. Massachussets Institute of Technology.

[Arndt et al. 2007] Richard Arndt, Rapha' el Troncy, Steffen Staab, Lynda Hardman, and Miroslav Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In 6th International Semantic Web Conference (ISWC), 2007.

[Avaro and Salembier 2001] Avaro, O.; Salembier, P.; "MPEG-7 Systems: overview," ***Circuits and Systems for Video Technology, IEEE Transactions on*** , vol.11, no.6, pp.760-764, Jun **2001**.

[Baader and Nutt 2002] F. Baader, W. Nutt. Basic Description LoRics. In the Description Logic Handbook, edited by F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider, Cambridge University Press, 2002, pages 47-100.

[Baader et al. 2007] Franz Baader, Ian Horrocks, and Ulrike Sattler. Description Logics. In Frank van Harmelen, Vladimir Lifschitz, and Bruce Porter, editors, ***Handbook of Knowledge Representation.*** Elsevier, 2007

[Bagdanov et al. 2007] Bagdanov, Andrew D.; Bertini, Marco; Bimbo, Alberto Del; Serra, Giuseppe; Torniai, Carlo. Semantic annotation and retrieval of video events using multimedia ontologies. International Conference on Semantic Computing, 2007. ICSC 2007. Volume , Issue , 17-19 Sept. 2007 Page(s):713 - 720

[Bateman et al. 2006] Bateman, S., Brooks and C., McCalla, G. (2006) Collaborative Tagging Approaches for Ontological Metadata in Adaptive ELearning Systems In the Proceedings of the Fourth International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL 2006) in conjunction with 2006 International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH2006), June 20, 2006. Dublin, Ireland.

[Batke et al. 2004] Jan-Mark Batke, Gunnar Eisenberg, Philipp Weishaupt, and Thomas Sikora. A Query by Humming system using MPEG-7 Descriptors. 116th Convention of Audio Engineering Society. May 2004. Berlin, Germany

[Baumann et al. 2002] S. Baumann, A. Kluter, and M. Norlien. Using natural language input and audio analysis for a human-oriented mir system. In Proceedings of the 2nd International Conference on Web Delivering of Music, Darmstadt, Germany, 2002.

[Bechhofer 2003] Sean Bechhofer. The DIG Description Logic Interface: DIG/1.1.February 7, 2003. Electronic Document. Available online from http://www.sts.tu-harburg.de/~r.f.moeller/racer/interfacel.l.pdf

[Berkan 2007] Riza C. Berkan. Semantic Search: An Antidote for Poor Relevancy. ReadWriteWeb.com archive, published on May 29, 2007. [online] Available at:

<http://www.readwriteweb.com/archives/semantic search antidote for poor relevancy.php> [Accessed 25 January 2009]

[Berners-Lee 2008] Berners-Lee T., 2008. Semantic Web Stack, [electronic print] Available at: http://www.w3.org/2000/Talks/1206-xml2k-tbl/slideIQ-0.html [Accessed 18 June 2009]

[Bertini et al. 2005] Bertini, M.; Cucchiara, R.; Del Bimbo, A.; Torniai, C. Video Annotation with Pictorially Enriched Ontologies. IEEE International Conference on Multimedia and Expo, 2005. ICME 2005. Volume , Issue , 6-6 July 2005 Page(s):1428 -1431

[Bertini et al. 2007] Bertini, Del Bimbo M;, Torniai A; C. (2007). Soccer video annotation using ontologies extended with visual prototypes. Proceedings of 2007 International Workshop on ContentBased Multimedia Indexing ***June 25-27, 2007 Bordeaux, France***

[Bertin-Mahiex et al. 2008] T. Bertin-Mahiex, D. Eck, F. Maillet, and P. Lamere. Autotagger: A model for predicting social tags from acoustic features on large music databases. Journal of New Music Reseach, 2008.

[Boley et al. 2001] Boley, H., Tabet, S., and Wagner, G. (2001). Design rationale of ruleml: A markup language for semantic web rules. The Semantic Web Working Symposium.

[Breitman and Prado 2003] Breitman, K.K.; do Prado Leite, J.C.S. Ontology as a requirements engineering product . 11th IEEE International Requirements Engineering Conference Proceedings.. Publication Year: 2003 , Page(s): 309-319

[Casey et al. 2008] Casey, M.A.; Veltkamp, R.; Goto, M.; Leman, M.; Rhodes, C.; Slaney, M. Content-Based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE. Volume 96, Issue 4. April 2008 Page(s):668 - 696.

[Ceccaroni 2001] Luigi Ceccaroni. Ontowedds- An Ontology-based Environmental Decision Support System for the Management of Waste Water Treatment Plants. PhD thesis. UNIVERSITAT POUT'ECNICA DE CATALUNYA, BARCELONA, SPAIN, OCTOBER (2001)

[Celma et al. 2006] Celma, O., Cano, P., & Herrera, P. (2006). Search sounds: An audio crawler focused on weblogs. In Proceedings of the international conference on music information retrieval (ISMIR).

[Chang and Amanda 2007] Chang, Yun-Ke; Spink, Amanda. 2007. Multimedia Chinese Web Search Engines: A Survey. Fourth International Conference on Information Technology, 2007. ITNG '07. 2-4 April 2007 Page(s):481 - 486.

[Cimiano and Handschuh 2003] Philipp Cimiano, Siegfried Handschuh. Ontology-based linguistic annotation. Proceedings of the ACL 2003 workshop on Linguistic annotation: getting the model right - Volume 19. July 2003

[Cognition 2008] Cognition, 2008. Cognition:Giving Technologies new meaning, [online] Available at: <http://wikipedia.cognition.com/info/who.html> [Accessed 15 January 2009]

[Connoly et al. 2004] Connolly, D.; Hendler, J.; Schreiber, G. 2004. Web-Ontology (WebOnt) Working Group, [online] Available at:< www.w3.org/2001/sw/WebQnt/> [Accessed 22 June 2008].

[Connor et al., 2005] Martin O'Connor, Holger Knublauch, Samson Tu, Benjamin Grosof, Mike Dean, William Grosso, Mark Musen. 2005. Supporting Rule System Interoperability on the Semantic Web using SWRL and Jess", Fourth International Semantic Web Conference (ISWC2005), Galway, Ireland 974-986, 2005.

[Connor, nd] Martin O'Connor. SWRLTab. [online] (Last updated 16 August 2008)Available at:< http://protege.cim3.net/cgi-bin/wiki.pl?SWRLTab>[Accessed 19 June 2009].

[Corthaut et al. 2008] Nik Corthaut, Sten Govaerts, Katrien Verbert, Erik Duval. CONNECTING THE DOTS: MUSIC METADATA GENERATION, SCHEMAS AND APPLICATIONS. *ISMIR 2008 – Session 2c – Knowledge Representation, Tags, Metadata.*

[Crysandt 2005] Crysandt, H. Hierarchical Sound Classification using Mpeg-7. 2005 IEEE 7th Workshop on Multimedia Signal Processing, Oct. 30 2005-Nov. 2 2005 Page(s):1 – 4.

[Damme et al. 2007] Van Damme, C., Hepp, M. & Siorpaes, K. (2007), FolksOntology, An Integrated Approach for Turning Folksonomies into Ontologies. *Proceedings of the ESWC 2007 Workshop "Bridging the Gap between Semantic Web and Web 2.0".*

[Day 2002] Neil Day. Search and Browsing. Chapter 20, Introduction to MPEG-7: Multimedia Content Description Interface. 2002. Published by John Willey and Sons Ltd.

[Dedek et al. 2008] J. Dedek, A. Eckhardt, L. Galambos, P. Vojtas. Uncertainty Reasoning for the World Wide Web. W3C Incubator Group Report 31 March 2008, available from http://www.w3.org/2005/Incubator/urw3/XGR-urw3/

[Delias et al. 2007] Delias, N.L.; Koutsoloukas, E.A.; Lioudakis, G.V.; Kaklamani, D.I.; Venieris, I.S. A Performance comparison of Ontolog Reasoning and Rule Engines. 16th IST Mobile and Wireless Communications Summit 2007. Pages 1-5

[Ding et al. 2007] Li Ding and Pranam Kolari and Zhongli Ding and Sasikanth Avancha, *Using Ontologies in the Semantic Web: A Survey,* in Ontologies: A Handbook of Principles` Concepts and Applications in Information Systems, Springer, pages 79-113, 2007.

[Donnadieu 2007] S. Donnadieu. “Mental Representation of the Timbre of Complex Sounds”, *Analysis, Synthesis, and Perception of Musical Sounds* (New York: Springer), 2007.

[Dori 2007] Dori, Y.J. (2007). Educational reform at MIT: Advancing and evaluating technology-based projects on- and off-campus. Journal of Science Education and Technology, 16(4)

[Eleftherohorinou et al. 2006] Eleftherohorinou, H., Zervaki, V., Gounaris, A., Papastathis, V., Kompatsiaris, Y., and Hobson, P.: "Towards a Common Multimedia. Ontology Framework"; Analysis of the Contributions to Call for a Common multimedia Ontology Framework Requirement; April, 2006. Available online: http://www.acemedia.org/aceMedia/files/multimedia_ontology/cfr/MM-Ontologies-Reqs-v1.3.pdf (last accessed: 01.04.2007)

[Fahad et al., 2008] Muhammad Fahad, Muhammad Abdul Qadir and Syed Adnan Hussain Shah. Evaluation of Ontologies and DL Reasoners. Intelligent Information Processing IV (IFIP International Federation for Information Processing). Volume 288, September 18, 2008, pages 17-27, Springer Boston

[Farooq et al., 2007] Farooq, A.; Shah, A.; Asif, K.H. Design of ontology in Semantic Web engineering process . HONET 2007. International Symposium on High Capacity Optical Networks and Enabling Technologies, 2007..Publication Year: 2007 , Page(s): 1 - 6

[Fellbaum 1998] Fellbaum, Ch. (ed.) (1998) "WordNet: An Electronic Lexical Database" MIT Press, Cambridge, MA.

[Fernandez et al. 2008] Fernandez, M.; Lopez, V.; Sabou, M.; Uren, V.; Vallet, D.; Motta, E.; Castells, P. Semantic Search Meets the Web. 2008 IEEE International Conference onSemantic Computing. 4-7 Aug. 2008 Page(s):253 – 260

[Fu et al. 2005] Xin Fu, Tom Ciszek, Gary Marchionini,, Paul Solomon. Annotating the Web: An exploratory study of Web users' needs for personal annotation tools. Proceedings of the

American Society for Information Science and Technology Volume 42, Issue 1, Date: 2005, Pages: NA

[Fudholi et al. 2009] Dhomas Hatta Fudholi, Noppadol Maneerat, Ruttikorn Varakulsiripunthl Yasushi Kato. Application of Protege, SWRL and SQWRL in Fuzzy Ontology-Based Menu Recommendation.2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2009) December 7-9, 2009

[Gang et al. 2008] Gang Chen; Tian-Jiang Wang; Herrera, P. A Novel Music Retrieval System with Relevance Feedback. 3rd International Conference on Innovative Computing Information and Control, 2008. ICICIC '08.18-20 June 2008 Page(s):158 - 158.

[Gangemi et al. 2006] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. Modelling Ontology Evaluation and Validation. ESWC 2006: 140-154

[Garcia and Celma 2005] Roberto Garcia and Oscar Celma. Semantic Integration and Retrieval of Multimedia Metadata. In 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot'05), Galway, Ireland (2005)

[Garshol 2004] Lars Marius Garshol: Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. J. Information Science 30(4): 378-391 (2004)

[Ghawi and Cullot 2009] Raji Ghawi, Nadine Cullot, "Building Ontologies from XML Data Sources," dexa, pp.480-484, 2009 20th International Workshop on Database and Expert Systems Application, 2009.

[Ghias et al. 1995] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: musical information retrieval in an audio database. In Proceedings of the 3rd ACM International Conference on Multimedia, San Francisco, CA, USA, 1995.

[Google 2008] Google, 2008. ***Google Technology Overview 2008.*** [online] Available at: <http://www.google.com/corporate/tech.html > [Accessed 4 June 2009]

[Goussevskaia et al. 2008] Goussevskaia, Olga; Kuhn, Michael; Lorenzi, Michael; Wattenhofer, Roger. From Web to Map: Exploring the World of Music. International Conference on Web Intelligence and Intelligent Agent Technology, 2008 IEEE/WIC/ACM Volume 1, 9-12 Dec. 2008 Page(s):242 - 248.

[Graybeal and Alexander 2009] Graybeal, J., Alexander, P. 2009. "What is an Ontology?." In The MMI Guides: Navigating the World of Marine Metadata, http://marinemetadata.org/guides/vocabs/ont/definition

[Gruber 1993] Gruber T. R. (1993) "A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5(2), 199-220.

[Gruber 1995] Gruber, T. R. (1995) "Toward principles for the design of ontologies used for knowledge sharing" in International Journal of Human-Computer Studies, Vol. 43, Issues 4-5, November 1995, pp. 907-928.

[Gruber 2008] Gruber, T. (2008). Collective knowledge systems: Where the Social Web meets the Semantic Web. Journal of Web Semantics 6(1): 4-13 (2008)

[Guarino and Giaretta 1995] Guarino N. and Giaretta P. (1995) "Ontologies and Knowledge Bases: Towards a Terminological Clarification" in N. J. I. Mars (ed.), Towards Very Large Knowledge Bases, IOS Press 1995, pp. 25-32.

[Hai et al. 2008] Hai Dong, Hussain, F.K.; Chang, E. A survey in semantic search technologies. 2nd IEEE International Conference on Digital Ecosystems and Technologies, 2008. DEST 2008. 26-29 Feb. 2008 Page(s):403 - 408.

[Hakia 2007] Hakia, 2007. Hakia Technology, [online] Available at: <http://company.hakia.com/technology.html> [Accessed 23 June 2008]

[Handschuh and Staab 2003a] S. Handschuh and S. Staab. Annotation for the Semantic Web. Volume 96 Frontiers in Artificial Intelligence and Applications. Edition: 2003.

[Handschuh and Staab 2003b] Siegfried Handschuh, Steffen Staab. CREAM: CREAting Metadata for the Semantic Web. Computer Networks 42 (2003) 579-598. Available at www.elsevier.com/locate/comnet

[Handschuh et al. 2001] Handschuh, S, Staab, S, and Maedche, A .CREAM - Creating relational metadata with a component-based, ontology-driven annotation framework. ACM K-CAP, 2001.

[Hardin, **2005]** Steve Hardin. **2005.** Tim Berners-Lee: The Semantic Web - Web of Machine-Processable Data. **Bulletin of the American Society for Information Science and Technology.** Volume 31. Issue 3. pages **12-13,** February/March **2005.** Willey online Library.

[Hardman 2005] L. Hardman. ***Canonical Processes of Media Production.*** In Proceedings of the ACM workshop on Multimedia for human communication. ACM Press, 2005.

[Hegaret, 2008] Philippe Le Hegaret, 2008. W3C Architecture Domain: Document Object Model Home. Availableat: <http://www.w3.org/DOM/Activitv.html > [Accessed 4 September 2008].

[Herman 2007] Herman I., 2007. Web Ontology Language (OWL), [online] Available at: <http://www.w3.org/2004/OWL/> [Accessed 14 April 2007]

[Herman nd] Herman I., nd. W3C Semantic Web Activity, [online] Available at: <www.w3.org/2001/sw/> [Accessed 22 June 2009].

[Hollink 2006] Laura Hollink. Semantic Annotation for Retrieval of Visual Resources. PhD thesis, 2006.

[Hollink et al. 2003] L. Hollink, A.Th. Schreiber, J. Wielemaker, and B. Wielinga. Semantic annotation of image collections. In Proceedings of the K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation, Florida, 2003.

[Horrocks et al. 2004] Horrocks I., Patel-Schneider P.F.,Boley H, Tabet S, Grosof B,Dean M.(2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission 21 May 2004. Electronic document. Available online from http://www.w3.org/Submission/SWRL/

[Horrocks et al., 2003] Ian Horrocks, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean. 2003. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. [online] Available at:< http://www.daml.org/2003/11/swrl/>[Accessed 17May 2007].

[Huang, 2008] Huang, T.; Li, W.; Yang, C. Comparison of Ontology Reasoners: Racer, Pellet, Fact++. American Geophysical Union, Fall Meeting, abstract #IN13A-1068, December 2008.

[Hunter 2003] Hunter, J. Enhancing the semantic interoperability of multimedia through a core ontology. IEEE Transactions on Circuits and Systems for Video Technology, Volume 13, Issue 1, Jan. 2003 Page(s):49 – 58

[Hunter and Little, 2007] Hunter J. and Little S., 2007. MPEG-7 Ontology. [online] Available at: <http://metadata.net/mpeg7/> [Accessed 15 February 2008]

[Hunter, 2000] Jason Hunter, 2000. JDOM- Open source API for XML Processing. [online] Available at: < http://www.jdom.org/> [Accessed on 20 April 2008]

[Kanzaki, 2003] Kanzaki, 2003. The Web KANZAKI. [online] (Last updated 31 October 2007) Available at: < http://www.kanzaki.com/ns/music.rdf> [Accessed on 20 April 2009]

[Kifer et al. 1995] Michael Kifer, Georg Lausen and James Wu. Logical Foundations of Object-Oriented and Frame-based Languages. Journal of ACM, 1995

[Kim et al. 2004] Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora. Audio Classification Based on MPEG-7 Spectral Basis Representations. IEEE Trans. Circuits Syst. Video Techn. 14(5): 716-725 (2004)

[Knees et al. 2006] Peter Knees , Markus Schedl , Tim Pohle , Gerhard Widmer, An innovative three-dimensional user interface for exploring music collections enriched, Proceedings of the 14th annual ACM international conference on Multimedia, October 23-27, 2006, Santa Barbara, CA, USA

[Knees et al. 2007] Peter Knees, Tim Pohle, Markus Schedl, and Gerhard Widmer. A Music Search Engine Built upon Audio-based and Web-based Similarity Measures. 8th International Conference on Music Information Retrieval (ISMIR 2007). September 2007, Vienna, Austria

[Koister 2009] Jari Koister. Applying Ontology, NLP and Semantic Search To Enterprise Collaboration. Article published on Semantic Universe on 02/03/09, http://www.semanticuniverse.com/articles-applving-ontology-nlp-and-semantic-search-enterprise-collaboration.html

[Koivunen 2005] Koivunen, M.-R. (2005). Annotea project. Electronic document. Accessed July 10, 2009. Available from http://www.annotea.org/eswc2005/01 koivunen final.pdf

[Kosch et al. 2005] Kosch, H.; Boszormenyi, L.; Doller, M.; Libsie, M.; Schojer, P.; Kofler, A. The life cycle of multimedia metadata. Multimedia, IEEE. Volume 12, Issue 1, Jan.-March 2005 Page(s):80 - 86

[Kostek 2003] Bozena Kostek. "Computing with words" Concept Applied to Musical Information Retrieval Electronic Notes in Theoretical Computer Science, Volume 82, Issue 4, March 2003, Pages 141-152

[Lamberti et al. 2009] Lamberti, F.; Sanna, A.; Demartini, C. A Relation-Based Page Rank Algorithm for Semantic Web Search Engines. IEEE Transactions onKnowledge and Data Engineering, Volume 21, Issue 1, Jan. 2009 Page(s):123 -136

[Lamere and Pampalk 2008] Paul Lamere and Elias Pampalk. Social Tags and Music Information Retrieval. ISMIR 2008

[Last.fm 2008] Last.fm, 2008. Last.fm Music Recommendation Service, [online] (Last updated 21 November 2008) Available at: < http://www.last.fm/about> [Accessed on 20 April 2009]

[Lenat 1995] Lenat, D. B. (1995) "A large-scale investment in knowledgeinfrastructure," in Communications of the ACM 38:11, 33-38.

[Lenat and Guha 1990] Lenat, D. B. and Guha R. V. (1990) "Building Large Knowledge-Based Systems", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

[Li and Ogihara 2006] Tao Li; Ogihara, M. Toward intelligent music information retrieval. IEEE Transactions on Multimedia. Volume 8, Issue 3. June 2006 Page(s):564 - 574.

[Li et al. 2007] Guan-yu Li; Sui-ming Yu; Sha-sha Dai. Ontology-Based Query System Design and Implementation. International Conference on Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP. 18-21 Sept. 2007 Page(s):1010 - 1015

[Lindsay et al. 2002] Adam T. Lindsay, Ian Burnett, Schuyler. Quackenbush and Melanie Jackson. Chapter 17, Introduction to MPEG-7: Multimedia Content Description Interface. 2002. Published by John Willey and Sons Ltd.

[Luo and Fan 2006] Luo, H. and Fan, J. 2006. Building concept ontology for medical video annotation. In ***Proceedings of the 14th Annual ACM international Conference on Multimedia*** (Santa Barbara, CA, USA, October 23 - 27, 2006). MULTIMEDIA '06. ACM, New York, NY, 57-60

[Luo et al., 2004] Jianli Luo; Jie Shen; Cuihua Xie; , "Segmenting the Web document with document object model," ***Services Computing, 2004. (SCC 2004). Proceedings. 2004 IEEE International Conference on*** , vol., no., pp. 449- 452,15-18 Sept. 2004

[MacManus 2007] Richard MacManus. Hakia Takes on Google withSemantic Web Technologies. March 23, 2007. [online] Available at: <http://www.readwriteweb.com/archives/hakia takes on google semantic search.php> [Accessed 11 September 2008].

[Maddage et al. 2006] N. C. Maddage, H. Li, and M. S. Kankanhalli. Music structure based vector space retrieval. In Proceedings of 29th Annual International ACM SIGIR Conference, Seattle, USA, 2006.

[Mahmoud 2005] Qusay H. Mahmoud, ***April 2005.*** Service-Oriented Architecture (SOA) and Web Services: The Road to Enterprise Application Integration, [online] Available at: <http://java.sun.com/developer/technicalArticles/WebServices/soa/> [Accessed 8 November 2008].

[Mandel and Ellis 2008] M. I. Mandel and D. P. W. Ellis. A Web-Based Gamefor Collecting Music Metadata. In Journal of New Music Research, 2008.

[March and Smith 1995] March S and Smith G. (1995) "Design and Natural Science Research on Information Technology". Decision Support Systems 15 (1995): 251-266.

[Martinez et al. 2002]Martinez, J.M.; Koenen, R.; Pereira, F. MPEG-7: the generic multimedia content description standard, part 1. IEEE Multimedia, Volume 9, Issue 2, April-June 2002 Page(s):78 - 87

[McCarthy 1980] McCarthy, J. (1980) "Circumscription - A Form of Non-Monotonic Reasoning", Artificial Intelligence, 5:13, 27-39.

[Midwinter 2007] Phill Midwinter. Is Google a Semantic Search Engine? Published on 26 March 2007. [online] Available at: <http://www.readwriteweb.com/archives/is_google_a_semantic_search_engine.php> [Accessed 10 September 2008].

[Miller 1995] Miller, G. A. (1995) "WordNet: A lexical database for English" in Communications of the ACM 38:11, 39-41.

[Miller and Swick 2003] Eric Miller and Ralph Swick. An Overview of W3C Semantic Web Activity. ***Bulletin of the American Society for Information Science and Technology—April/May 2003.***

[Misdariis et al., 2010] Nicolas Misdariis, Antoine Minard. Patrick Susini. Guillaume Lemaitre. Stephen McAdams, and Etienne Parizet. Environmental Sound Perception: Metadescription and Modeling Based on Independent Primary Studies. EURASIP Journal on Audio, Speech, and Music Processing, Volume 2010 (2010), Article ID 362013, 26 pages

[Mitrovic et al. 2007] Mitrovic, Dalibor; Zeppelzauer, Matthias; Eidenberger, Horst. Analysis of the data quality of audio descriptions of environmental sounds. Research Article from: Journal of Digital Information Management, April 1, (2007) [Mostowfi et al. 2005] Farhad Mostowfi, Farshad Fotouhi & Anthony Aristar. Ontogloss: An Ontology-Based Annotation Tool. Proceedings of the 2005 EMELD workshop on Linguistic Ontologies and Data Categories for Language Resources. July 2005.

[MPEG-7 Audio Encoder 2008] MPEG-7 Audio Encoder, 2008. MPEG-7 Audio Encoder from Technical University of Berlin, [online] Available at: <http://mpeg7lld.nue.tu-berlin.de/> [Accessed 4 June 2008]

[MPEG-7 Part 5 2003] MPEG-7 Part 5, 2003. Multimedia Content Description Interface Part 5. ISO/IEC 15938-5:2003

[Naphade et al. 2006] Milind Naphade, John R. Smith, Jelena Tesic, Shih-Fu Chang, Winston Hsu, Lyndon Kennedy, Alexander Hauptmann, Jon Curtis.: Large-Scale Concept Ontology for Multimedia. ***IEEE MultiMedia,*** vol. 13, no. 3, pp. 86-91, July-Sept. (2006)

[Oneill, 2009] Dan ONeill, 2009. ID3. [online] Available at: <http://www.id3.org/> [Accessed 30 October 2010]

[Open University 2007a] Open University, 2007.The Technology of Music: Investigating sound. 2007. The Open University, UK.

[Open University 2007b] Open University, 2007. Technology of Music: Prelude. 2007. The Open University, UK.

[Ossenbruggen et al. 2004] Ossenbruggen, J. v., Nack, F., and Hardman, L. 2004. That Obscure Object of Desire: Multimedia Metadata on the Web, Part 1. ***IEEE MultiMedia*** 11, 4 (Oct. 2004), 38-48

[Ossenbruggen et al. 2005] Nack, F.; van Ossenbruggen, J.; Hardman, L. That obscure object of desire: multimedia metadata on the Web, part 2. IEEE Multimedia, IEEE Volume 12, Issue 1, Jan.-March 2005 Page(s):54 - 63

[Pampalk and Goto 2007] Elias Pampalk and Masataka Goto.(2007). MusicSun: A New Approach to Artist Recommendation. 8th International Conference on Music Information Retrieval (ISMIR 2007).September 23rd-27th 2007, Vienna, Austria.

[Pampalk et al. 2006] Elias Pampalk, Masataka Goto: MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling . 7th International Conference on Music Information Retrieval (ISMIR 2006), pp . 367-370. Victoria, Canada, October 8-12, 2006.

[Pan et al. 2006] Jeff Z. Pan, Giorgos Stoilos, Giorgos B. Stamou, Vassilis Tzouvaras, and Ian Horrocks. f-SWRL: A Fuzzy Extension of SWRL. ***J. on Data Semantics,*** VL28-46, 2006

[Parekh et al. 2004] Viral Parekh, Jin-Ping Gwo and Tim Finin. Ontology based Semantic Metadata for Geo-Science Data. The International MultiConference in Computer Science and Computer Engineering, June 2004.

[Passin 2004] Passin, Thomas B. Explorer's Guide to the Semantic Web. Chapter 4. Greenwich Manning Publications, 2004.

[Perez 2009] Juan Carlos Perez. Google Rolls out Semantic Search Capabilities. PCWorld Business Center. Tuesday, March 24, 2009. [online] Available at: <http://www.pcworld.com/businesscenter/article/161869/Roogle rolls out semantic sea rch capabilities.html> [Accessed 5 May 2009].

[Petridis et al. 2006a] Petridis, K.; Bloehdorn, S.; Saathoff, C.; Simou, N.; Dasiopoulou, S.; Tzouvaras, V.; Handschuh, S.; Avrithis, Y.; Kompatsiaris, Y.; Staab, S. Knowledge representation and semantic annotation of multimedia content. Vision. Image and Signal Processing. IEE Proceedings- Volume 153, Issue 3. 8 June 2006 Page(s):255 - 262

[Petridis et al. 2006b] Kosmas Petridis, Dionysios Anastasopoulos, Carsten Saathoff, Norman Timmermann, Yiannis Kompatsiaris, and Steffen Staab. M-OntoMat-Annotizer: Image Annotation Linking Ontologies and Multimedia Low-Level Features. Knowledge-Based Intelligent Information and Engineering Systems. Lecture Notes in Computer Science. Volume 4253/2006, Springer Berlin / Heidelberg, pp 633- 640

[Pidcock 2003] Woody Pidcock. What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model? January 15, 2003. Available online at http://www.metamodel.com/article.php?storv=20030115211223271; accessed on December 18, 2007

[Powerset 2008] Powerset, 2008. Frequently Asked Questions, [online] Available at: <http://www.powerset.com/faq/> [Accessed 10 July 2009]

[Raimond et al. 2007] Yves Raimond, Samer Abdallah, Mark Sandler.THE MUSIC ONTOLOGY. 2007. Austrian Computer Society

[Reeve and Han 2005] Lawrence Reeve, Hyoil Han. Web technologies and applications (WTA): Survey of semantic annotation platforms. Proceedings of the 2005 ACM symposium on Applied computing SAC '05. March 2005.

[Rho et al. 2009] Seungmin Rho, Seheon Song, Eenjun Hwang, and Minkoo Kim. COMUS: Ontological and Rule-Based Reasoning for Music Recommendation System. Third International Conference on Digital Society, 2009. ICDS '09. February 2009. page(s): 304-309

[Ruvane 2006] Mary B. Ruvane. Defining annotations: a visual (re)interpretation. Proceedings of the American Society for Information Science and Technology. Volume 43, Issue 1, Date: 2006, Pages: 226

[Ruxanda et al. 2008]Ruxanda, M.M.; Nanopoulos, A.; Jensen, C.S.; Manolopoulos, Y. Ranking music data by relevance and importance. 2008 IEEE International Conference on Multimedia and Expo. June 23 2008-April 26 2008 Page(s):549 - 552.

[Saathoff et al. 2006] Carsten Saathoff, Simon Schenk, Steffen Staab. Requirements for a Multimedia Ontology Framework. 3rd European Semantic Web Conference, 11-14 June 2006. Montenegro.

[Salembier 2002] Phillippe Salembier and John R Smith. Overview of Description Schemes and Schema Tools. Chapter 6, Introduction to MPEG-7: Multimedia Content Descrioption Interface. 2002. Published by John Willey and Sons Ltd.

[Sarkar et al. 2007] M. Sarkar, B. Vercoe, and Y. Yang, "Words that describe timbre: a study of auditory perception through language". Language and Music as Cognitive Systems Conference (LMCS-2007), Cambridge, UK, May 11-13, 2007.

[Schmidt-Jones, 2010] Schmidt-Jones C., 2010. Melody - Connexions module, [online] Available at: <http://cnx.orR/content/mII647/latest/> [Accessed 30 April 2010]

[Schneider and Horrocks 2006] Peter Patel-Schneider and Ian Horrocks. OWL 1.1 Web Ontology Language Overview. W3C Member Submission, 19 December 2006. [online] Available at <http://www.w3.orR/Submission/owlII-overviewMAccessed 30 March 2008]

[Schroeter et al. 2006] Schroeter, R., Hunter, J., Guerin, J., Khan, I., and Henderson, M. 2006. A Synchronous Multimedia Annotation System for Secure Collaboratories. In ***Proceedings of the Second IEEE international Conference on E-Science and Grid Computing*** (December 04 - 06, 2006). E-SCIENCE. IEEE Computer Society, Washington, DC.

[Seth 2007] Amit Sheth. Semantic Meta Data for Enterprise Information Integration. DM Review Magazine, July 2003. [online] Available at <http://www.dmreview.com/issues/20030701/6962-l.html> [Accessed 25 October 2007].

[Shirky 2005] Clay Shirky. 2005. "Ontology Is Overrated". O'Reilly ETech conference. March 2005.

[Simon 1969] Simon, Herbert Alexander (1969) (in English). ***The sciences of the artificial.*** Karl Taylor Compton lectures. Cambridge: M.I.T. Press.

[Smith 2004] Smith B. (2004) "Beyond concepts: Ontology as reality representation" in Proceedings of the Third International Conference on Formal Ontology in Information Systems (FOIS 2004). Amsterdam: IOS Press; 2004, p.73-84.

[Smith and Schirling 2006] Smith, J.R.; Schirling, P. Metadata standards roundup. Multimedia, IEEE, Volume 13, Issue 2, April-June 2006 Page(s):84 - 88

[Smith and Welty 2001] Smith, B. and Welty, Ch. (2001) "Ontology: Towards a New Synthesis" in Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS2001). ACM Press, 2001.

[SNOEK et al. 2006] CEES G. M. SNOEK and MARCEL WORRING, ALEXANDER G. HAUPTMANN. Learning Rich Semantics from News Video Archives by Style Analysis.ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 2, No. 2, May 2006, Pages 91-108.

[Sonera 2003] Sonera White Paper, 2003. The MPEG-7 White Paper from Sonera Media Lab. [online] Finland: Telisonera (Published 2003) Available at <http://www.medialab.sonera.fi/workspace/MPEG7WhitePaper.pdf> [Accessed 17 March 2007].

[Sowa 1984] Sowa, J. F. (1984) "Conceptual Structures. Information Processing in Mind and Machine", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA

[Sowa 2000] John F. Sowa. ***Knowledge Representation: Logical, Philosophical, and Computational Foundation.*** Pacific Grove, CA: Brooks/Cole, 2000

[Sowa 2006] Sowa, J. F. (2006) "A Dynamic Theory of Ontology" in B. Bennett & C. Fellbaum, eds., Formal Ontology in Information Systems, IOS Press, Amsterdam, 2006, pp. 204-213.

[Staab, 2006] Steffen Staab, 2006. Semantics in Multimedia Content for Multimedia Use,University of Koblenz - Landau, Germay. ***Workshop:*** Knowledge in Multimedia Content 1ST Event 2006. [online] Available at <http://mklab.iti.gr/ist2006> [Accessed 20 July 2008].

[Stefik and Conway 1982] Stefik, M. and Conway L. (1982) "Towards the Principled Engineering of Knowledge", AI Magazine, 3(3): 4-16.

[Stevens, 2001] Stevens Robert, 2001. Knowledge representation languages, [online] Available at: <http://www.cs.man.ac.uk/~stevensr/onto/nodel4.html > [Accessed 10 January 2008]

[Storr 1997] Anthony Storr. Music and the Mind. HarperCollins Publishers, London, 1997

[Suh and Bederson 2007] Bongwon Suh and Benjamin B. Bederson. Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition. ***Interacting with Computers***, ***Volume 19, Issue*** 4, ***July 2007, Pages 524-544***

[Suyoto et al. 2008] Suyoto, I.S.H.; Uitdenbogerd, A.L.; Scholer, F. Searching Musical Audio Using Symbolic Queries. Audio, Speech, and Language Processing, IEEE Transactions on Volume 16, Issue 2, Feb. 2008 Page(s):372 - 381.

[Tsekeridou et al. 2006] Sofia Tsekeridou, Athina Kokonozi, Kostas Stavroglou, Christodoulos Chamzas: MPEG-7 Based Music Metadata Extensions for Traditional Greek Music Retrieval. MRCS 2006: 426-433.

[Tsinaraki et al. 2004] Tsinaraki C., Polydoros P., Christodoulakis S. Integration of OWL ontologies in MPEG-7 and TV-Anytime compliant Semantic Indexing. In the proceedings of the 16th International Conference on Advanced Information Systems Engineering (CAISE), pp. 398-413, Riga, Latvia, June 2004

[Tsinaraki, 2007] Tsinaraki, Chrisa; Polydoros, Panagiotis; Christodoulakis, Stavros. Interoperability support between MPEG-7/21 and OWL in DS-MIRF. *IEEE Transactions on Knowledge and Data Engineering*, v 19, n 2, February, 2007, p 219-232

[Turnbull et al. 2007a] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Towards musical query-by-semantic description using the CAL500 data set. SIGIR 2007.

[Turnbull et al. 2007b] Douglas Turnbull, Ruoran Liu, Luke Barrington, Gert Lanckriet. A Game-Based Approach for Collecting Semantic Annotations of Music. 8th International Conference on Music Information Retrieval (ISMIR 2007).September 23-27 2007, Vienna, Austria.

[Turnbull et al. 2008] D. Turnbull, L. Barrington, G. Lanckriet. *Five Approaches to Collecting Tags for Music.* ISMIR 2008.

[Uren et al. 2006] Uren, Victoria; Cimiano, Philipp; Iria, Jose; Handschuh, Siegfried; Vargas-Vera, Maria; Motta, Enrico; Ciravegna, Fabio. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Web Semantics*, v 4, n 1, January, 2006, p 14-28.

[Vinet 2004] Hugues Vinet. The Representation Levels of Music Information. LNCS, Computer Music Modelling and Retrieval Springer Volume 2771/ pages 203-216 (2004).

[Weinberger 2007] D. Weinberger. *How tagging changes peoples relationship to information and each other. Pew Internet & American Life Project*. (2007)

[Weller 2007] K. Weller. *Folksonomies and Ontologies. 2007. Two New Players in Indexing and Knowledge Representation.* In H. Jezzard (Ed.) Applying Web 2.0: Innovation, Impact and Implementation. Proceedings of the Online Information Conference 2007, London, Great Britain (pp. 108-115). London: Learned Information Europe Ltd.

[Welty et al. 1999] Welty, C., Lehmann, F., Gruninger, G., and Uschold, M. (1999) "Ontology: Expert Systems All Over Again?" Invited panel at AAAI-99: The National Conference on Artificial Intelligence. Austin, Texas.

[Whitman 2002] Whitman, Brian and Steve Lawrence (2002). Inferring Descriptions and Similarity for Music from Community Metadata. In Voices of Nature. Proceedings of the 2002 International Computer Music Conference. pp 591-598. 16-21 September 2002, Göteborg, Sweden.

[Whitman 2004] Whitman, Brian. Whitman PhD thesis. Learning the meaning of music. MIT, April 14 2004.

[Whitman and Lawrence 2002] B. Whitman and S. Lawrence. Inferring Descriptions and Similarity for Music from Community Metadata. In Proceedings of the 2002 International Computer Music Conference, Gothenburg, Sweden, 2002.

[Wielemaker et al. 2001] J.Wielemaker A.Th. Schreiber, B. Dubbeldam and B.J. Wielinga. Ontology-based photo annotation. IEEE Intelligent Systems, May/June 2001.

[Wikipedia 2008] Wikipedia, 2008. Knowledge Representation and Reasoning, [online] Available at: <http://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning > [Accessed 10 January 2008].

[William 2006] William M.K., 2006. Research Methods Knowledgebase, [online] Available at: <http://www.socialresearchmethods.net/kb/index.php/> [Accessed 7 April 2008]

[Yang 2007] Yang, S.-Y., An ontological website models-supported search agent for web services, Expert Systems with Applications (2007), doi:10.1016/j.eswa.2007.09.024

[Yokoi 1995] Yokoi, T. (1995) "The EDR electronic dictionary" in Communications of the ACM 38:11, 42-44.

[Zhang 2007] Zhang, J. (2007). Ontology and the Semantic Web. In Tennis, Joseph T., Eds. Proceedings North American Symposium on Knowledge Organization 2007 1, pages pp. 9-20, Toronto, Ontario.

[Zou et al 2008] Zou; Bofeng Zhang; Yanglan Gan; Jianwen Zhang. An Ontology-Based Methodology for Semantic Expansion Search. Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Volume 5, 18-20 Oct. 2008 Page(s):453 -457

Appendix A

Definition of Musicological terms[24]

Music unfolds in time and hence it has got tempo or duration. Complexities of human perception of music are related to the temporal aspects such as beat, rhythm and tempo. Again, music is conventionally described using terms like melody, harmony, rhythm, and dynamics [146]. The composer utilizes the tools of composition for the intimacies of musical elements - melody, harmony, rhythm, and dynamics – all together the pattern we hear is a known as song. Rhythm blends the expression of harmony & dynamics with the tempo based on melody that forms the soul of the song.

Tempo

Musical notes unfold in time. Music has got duration or in other word tempo. Complexities of human performance and perception of music is related to the temporal aspect of music in several ways. The start and end of musical notes are related to the underlying beat. Music conventionally is described using temporal aspects such as beat, rhythm and tempo.

Beat

Pulse or Beats are regularly spaced instantaneous temporal markers to which musical events are related. The word beat also refers to the time interval between two consecutive temporal markers. In most cases, audible repetitive events coincide with the beat whereas sometimes the beat may be inaudible.

The duration of notes are conceived in terms of the underlying beat rate e.g. for a certain number of beats or for a fraction of a beat (subdivision). There are standard patterns of subdivisions and commonly used patterns are based on simple numerical ratios and they have characteristic sounds. However, musical parts are not always regular.

When presented with a sequence of identical equally spaced beats, people generally hear them as unequally accented e.g. the ticking of a clock. Though the ticks of a clock are all equal

[24] Technology of Music: Prelude. 2007. The Open University, UK

the listener might interpret the sound as a recurring two beat cycle: "tick, tock" (beat pattern – 1, 2, 1, 2, 1,2) or might imagine as a three beat cycle: "tick, tock, tock" (1 2 3 1 2 3 1 2 3) or as a four beat cycle with pattern like 1 2 3 4 1 2 3 4 etc. The common thing about these patterns are that beat 1 is the strongest of each cycle i.e. beat 1 ha i.e. beat 1 has got the strongest accentuation. Groups of evenly spaced beats appear to be set off from one another by equally spaced accented beats. Regular beats fall into repeating cycles of strong and weak accentuation. Each cycle begins with a strong accent. A cycle of accentuation is called a bar.

Time Signature

The time signature is a way of indicating how many beats there are in a bar and what note value represents the beat in terms of time interval between two beats.

Harmonic

A harmonically related series of frequencies has the following pattern F1, 2f1, 3f1, 4f1 ....

Where f1 is called the fundamental frequency. Sine waves with harmonically related frequencies are called harmonics. Harmonics are numbered, the first harmonic has the frequency f1, the second harmonic has the frequency 2f1 and so on.

Harmony is the relation of notes to notes and chords to chords as they are played simultaneously. Harmonic "patterns" are established from notes and chords in successive order.

Rhythm

Rhythm is the pattern of note values used in any section of the music, usually together with their meter. The concept of rhythm does not apply to just one bar's worth of note values. Rhythm means musical time. As meter regulates and pulsates a poem, rhythm organizes music in much the same way. The regular pulsations of the music are called the beat. Time patterns in music are referred to in terms of meter. When the melody falls on notes that occur between beats, it is said to be syncopated time. Along with rhythm comes the idea of rate or pace. Not every song is slow or fast. Tempo is the musical term that indicates the overall pace of an arrangement e.g. grave & slow or fast & cheerful etc.

Dynamics

Without dynamics, music lacks the emotion behind the musical thought. Dynamics tell the performer when to play loudly or more softly and when to change from one to the other. From pianissimo (as soft as you can play) to fortissimo (the loudest you can play), music ranges from a whisper to the fullest of sound.

Melody

Melody is a successive line of tones or pitches that are characterised by range, shape, movement or motion. Melody is structured by its length & intensity including cadence (final ending to a musical section) and climax (high point of intensity). Melodic intervals are those that are linear and occur in sequence, while harmonic intervals are sounded at the same time. Whether or not a harmony is pleasing is a matter of personal taste, as there are consonant and dissonant harmonies, both of which are pleasing to the ears of some and not others.

Timbre

Timbre is the characteristic sound of an instrument or voice that enables to differentiate the sound from other instruments when they are playing the same note. Sound produced by conventional musical instrument or voices are usually not sinusoidal rather they are mixture of sine waves. This non-sinusoidal character is related to the *timbre*.

Appendix B

We tried to find the song "Lucky Star by Madonna but faster" mentioning tempo (speed) of the song in the search query using several search engines e.g. Google, Yahoo, Track9, AllMusicGuide, Last.fm etc. But none ofthe were able to find us the exact musical object we looked for. Below are the screen shots.

Google Search Results for "Lucky Star by Madonna but faster"

Lucky Star by Madonna but foster ■Google Starch - Windows Internet Exploie

031.ro.cfc/seaich^hi=er.gc;lz=IW. AD3r_er-<.sBac=Luckv-K ar+by+Madonna+but+faster&meta=&aq=fSaoq=

File Edit View Favorites Tools Help

Go gie Lucky Star by Madonna but f t § Share - §3 - 5xJewikj - ^ Check' T Lucky ^5tar by Madonna ^but [C^ faster ^ '

Lucky Star by Madonna but faster - Gxtgle Search ft 1 r/n - .-/Page ' , l

up lo date. Learn More Share my location j | Don't share | 0 Remember for This site

Web Imaaes Videos Maps News Shopping Mail mote i Web History I Search settings I StflLin

Google Lucky Star by Madonna but faster

Search: the web 0 pages from the UK

Web S Show options,.. Results 1 -10 of about 61,000 for Lucky Star by Madonna but faster. (0.24 seconde)

Madonna Song Lvncs!
Fancy cars that go s'ery fast you know they never last, no no. What you need is a big strong hand .... You may be my lucky star But lim the luckiest by far...
www.amandashome.cGmi'madonna.html - Cached

Madonna Song Lyrics Quiz
Lucky Star 5 And when the music starts I never wanna stop, it's gonna drive ... Faster than the speeding light she's flying, trying to remember where it all began ... If I'm smart then I'll run away, but I'm not so I guess ffl stay ...
www.musicquizworld.com >Music Quizzes - Cached - Similar

Absolute Madonna » The Immaculate Collection
I post - Last post: 19 Jul 2009
Lucky Star Borderline Like A Virgin Material Girl Crazy For You .... The Immaculate Collection contains the bulk of Madonna's hits but there are ... Furthermore, several songs are faster than their original versions and ...
ab30lulemadoniiQ.com/^page_id=35t2 - Cached - Similar

Pop diva Madonna planning babv with Jesus Luz I CaicuttaTube
5. lucky star [1984 remix edit] (1984) – 'Idont know but this version ismuch .... "Hung Up" is a fast and catchy track that you will love because Madonna ...
calcuttatube.com/pop-diva-inadonna baby.../52251/ - United States - Cached

Madonna - The Immaculate Collection R@Pmp3 320h33tschon55 Torrent...
Download 5x Faster ... Although Madonna re-recorded the vocals for the song "Lucky Star" for the compilation, all ofthe original vocals on ... There is no definite Madonna compilation yet but we may have to wait a few years for that. ...
www.torrentdcvvnloads.net >Categones >Other >Other - Cached

Madonna. Hard Candy I Track Bv Track. Review - P.Viktor
II May 2008 ... Heartbeat I This song opens up like a modern-day LuckyStar especially on albums dominated by fast tempo songs but I think this is a ...
pviktor.co.uk/p_viktor_/2008/05/inodonna-hard-ca.html - Cached

[10] Using natural language input and audio analysis for a human ...
File Format PDF/Adobe Acrobat - Quick View
by S Baumann - Cited by 16 - Related articles
Simplified, a possible query "like lucky star by Madonna, but faster is processed as follows Initiated hu thp Custnmpr Fnvirnnmnnt and Ssmhp and handed ...

S 'uoogle Mad - .«ust t : 1 or.

Yahoo Music Search Engine for "Lucky Star by Madonna but faster"

0 •• :© f http://search.music.yahoo.com/searchyhri=all8ip=IiJcky+StartJ-by-t-Madorina-"-but+faster v X P -

Rle Edt View Favorites Tools Hdp

G o glc mu»c so»ch engne ~ Search *«* J5» * * g j Share* £££) - Sriewtki - ^ Check- aa) Translate * AutoFil * |Ct muse search [C^ engne Sgn In *

© ! Lucky Start by Madonna but Paster - Yahoo1 Music Se... ft} w - [• • ^ tool.

New User? Sign Up I Sign In I Help **Upgrade to Safer IE8** © ? Yahoo? ■ Mail

Y * M U S IC

6 MUSIC videos * Radio * Downloads'- Artists * Exclusives* Blogs* My Music*

Search Yahoo! Music all **w (Search ^ill.'ujx**

shaadi".C U m **Mjtiimoni.il** Ge-iuice

If you **didn't** find **what you're looking for, try another** search.

Looking for Partner **Profiles**
from your **community?**
Start Now i

Rhapsody Unlimited. Listen to millions of sonos without paying per track - try it free!
Yahoo! Music FoxyTunes. Control your music while you surf Now available with Firefox 3
Yahoo! Music Backstage. The hottest artists, charts, and videos
Afld top songs, albums, and videos to My Yanool and RSS

Yahoo! Music Worldwide f * T |

Pnv.cyPe.cy | AocutourA.. | i.rm, o.s.rv*. | Cccynjr,® Poucy | F. MMOK | H«

Copyright © 2010 Yahoo1 Inc. All Rights Reserved J M / MUSIC

Q) **Int.™.** *. **100%** ' j

Track9.com for "Lucky Star by Madonna but faster"

(» 7 ; http://www track9.com",5earcn-LUcky-S ar-by-Madoma-but-faster&page« 1M;type«mp3 * ** x I * -

Rle Edt View Favorites Tools Hdp

G o glo music search engne Search • * c ^ * Q Share * - Sidewikj - ^ Check - #*j Translate * AutoFil * [C^ mu- c search [5^ engne ^ * 5gn In -

Jt ■&' C Star by Madonna but faster - Track9.co ftl - mo " lli> * J Tools *

T R A C K 9 . C O M

Lucky Star by Madonr Search:

i B c c .r: 5 jv

FREE MUSIC

PLAY

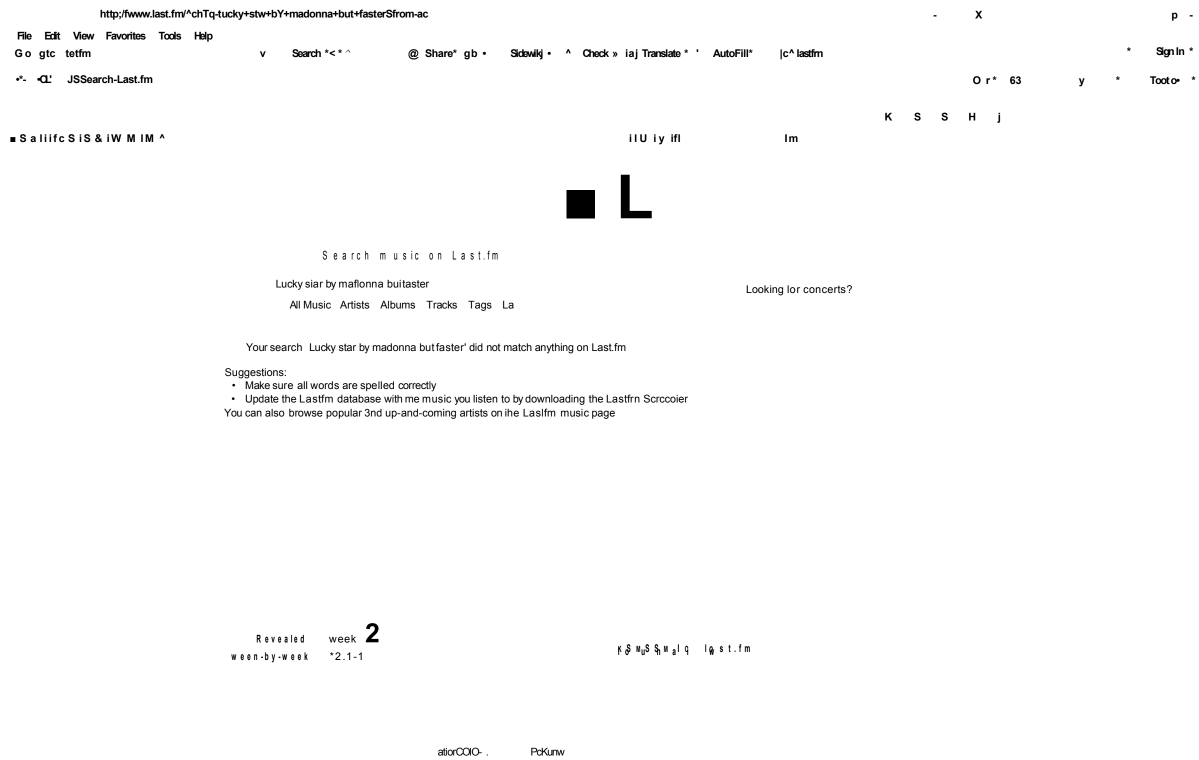
•JKgB8tf-O00'S

Rating

Playtintei

tat.oriCUIU- '2J I >»:un»ntc -Miciosof. . r Gf.ugle Mail ■ £ **Lucky Star by Madcm.**

Last.fm for "Lucky Star by Madonna but faster"



AIIMusic Guide also returned empty results.

Appendix C

Sample MPEG-7 Feature Extractor Output


```
<?xml version="1.0" encoding="iso-8859-1"?>
<!-- TU Berlin Audio Analyzer v1.0: http://www.nue.tu-
berlin.de/forschung/projekte/mpeg7/ -->
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
      xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 Mpeg7-2001.xsd">
<Description xsi:type="ContentEntityType">
<MultimediaContent xsi:type="AudioType">
<Audio xsi:type="AudioSegmentType">
.
.
.
<AudioDescriptor xsi:type="LogAttackTimeType">
            <Scalar>  -0.420216</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="TemporalCentroidType">
            <Scalar>  4.841836</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="SpectralCentroidType" >
            <Scalar>  628.982727</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="HarmonicSpectralCentroidType">
            <Scalar>  600.172363</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="HarmonicSpectralDeviationType">
            <Scalar>  -0.091171</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="HarmonicSpectralSpreadType">
            <Scalar>  0.596429</Scalar>
</AudioDescriptor>

<AudioDescriptor xsi:type="HarmonicSpectralVariationType">
            <Scalar>  0.225479</Scalar>
</AudioDescriptor>
...
....
<AudioDescriptor xsi:type="AudioFundamentalFrequencyType"
loLimit="25.0" hiLimit="16000.0">
      <SeriesOfScalar hopSize="PT10N1000F" totalNumOfSamples="10">
            <Raw>  595.945923 84.482758 117.914436 117.599998
45.370369 117.914436 73.745819 73.745819 197.757843 42.241379
      </Raw>
            <Weight> 0.949303 0.949303 0.949303 0.949303 0.949303
0.949303 0.949303 0.949303 0.949303 0.949303
      </Weight>
      </SeriesOfScalar>
</AudioDescriptor>
</Audio>
</MultimediaContent>
</Description>
</Mpeg7>
```

Appendix D

Snapshot of MPEG-7 Audio encoding tool from Technical University of Berlin

li^j » «^ http://webcache.googteusercontent.conysearch'q-cache:http://mpeg7ld.niJ8.tu-berin.def *] V x __|! f/ ■

Fite Edit View Favorites Took Help

Go gle technical university of Detin v Search • - c j - @ Share- ^ - Sidewto • Cheek- $aj Translate - 'iJAutoFil- ^ technical jC^ umversty ^ of ^ berlr ^ * Sign In -

•^TU-Berin: MPEG-7 Audb Analyzer -LowLevelDesdp... * Lj AW' - J Tools -

*

This is Google's cache of hltpT/mpegTild.nue.tu-frerlm.cte/- If is a snapshot ofthe page as it appeared on 30 Jan 2010 13 58 22 GMT. The current page could have changed in the meantime Leam more

Text-onlv version

Technical University of Berlin
Comrtrumcahon Systems Department
Project: MPEG-7 Annotation of Video Sequences

0 0 §Z7C° **MPEG-7 Audio Analyzer**
dBsin **Low Level Descriptors Extractor**

[*Home* | Upload | Choose descriptors | Receive the results]

Introduction

The audio Low Level Descriptors (LLDs) are a set of sound features defined within the new MPEG-7 standard. They measure several characteristics of sound, which are then stored as an XML file that serves as a compact representation of the analyzed audio The LLDs are the basis to create advanced MPEG-7 audio content-based applications

The TU Berlin Audio Analyzer implements all 17 audio descriptors defined in the standard Here, it is possible to compute LLDs of an audio file of your choice and receive the chosen LLDs in an MPEG-7 XML file

Steps

1. Send us your audio file m WAV or MP3 file format.
2. Choose the MPEG-7 Audio Low Level Descriptors (LLD) and specify' the parameters on the selected descriptors.
3. Receive the XML encoded results of the calculation.

Select the Audio File

Send us a WAV or MP3 audio file with the following properties:

- the file size has to be less than 1 MByte for WAV and less than 300 KByte for MP3.
- the audio file has to contain only one audio channel

j| Browse... [

Done Internet +,, 100% -

iEE5S52BIB23fli a W&m S i

Appendix E

The OWL1.0 encoded representation of the ***mpeg-7Music*** Ontology


```xml
<?xml version="1.0"?>
<rdf:RDF
    xmlns="http://www.owl-ontologies.com/Ontology1268728302.owl#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
    xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:sqwrl="http://sqwrl.stanford.edu/ontologies/built-
ins/3.4/sqwrl.owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:swrl="http://www.w3.org/2003/11/swrl#"
    xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:swrla="http://swrl.stanford.edu/ontologies/3.3/swrla.owl#"
  xml:base="http://www.owl-ontologies.com/Ontology1268728302.owl">
  <owl:Ontology rdf:about="">
    <owl:imports
rdf:resource="http://swrl.stanford.edu/ontologies/3.3/swrla.owl"/>
    <owl:imports
rdf:resource="http://sqwrl.stanford.edu/ontologies/built-
ins/3.4/sqwrl.owl"/>
  </owl:Ontology>
  <owl:Class rdf:ID="mp3">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="ResourceFormat"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="TimeLine"/>
  <owl:Class rdf:ID="MelodicExpressions">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="MusicalConcepts"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="Motion">
    <rdfs:subClassOf rdf:resource="#MelodicExpressions"/>
  </owl:Class>
  <owl:Class rdf:ID="Brightness">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="TimbralExpressions"/>
    </rdfs:subClassOf>
  </owl:Class>
  <owl:Class rdf:ID="MusicSegment"/>
  <owl:Class rdf:about="#TimbralExpressions">
    <rdfs:subClassOf rdf:resource="#MusicalConcepts"/>
  </owl:Class>
  <owl:Class rdf:ID="MediaTime">
    <rdfs:subClassOf rdf:resource="#TimeLine"/>
  </owl:Class>
  <owl:Class rdf:ID="Sharpness">
    <rdfs:subClassOf rdf:resource="#TimbralExpressions"/>
  </owl:Class>
  <owl:Class rdf:ID="Shape">
    <rdfs:subClassOf rdf:resource="#MelodicExpressions"/>
  </owl:Class>
  <owl:Class rdf:ID="PhysicalTime">
    <rdfs:subClassOf rdf:resource="#TimeLine"/>
  </owl:Class>
  <owl:Class rdf:ID="wav">
    <rdfs:subClassOf rdf:resource="#ResourceFormat"/>
```

```
  </owl:Class>
  <owl:ObjectProperty rdf:ID="correspondsTo"/>
  <owl:ObjectProperty rdf:ID="denotedBy">
    <rdfs:range rdf:resource="#TimeLine"/>
    <rdfs:domain rdf:resource="#MusicSegment"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="describedBy">
    <rdfs:domain rdf:resource="#MusicSegment"/>
    <rdfs:range rdf:resource="#MusicalConcepts"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="atMediaTime">
    <rdfs:range rdf:resource="#MediaTime"/>
    <rdfs:subPropertyOf>
      <owl:ObjectProperty rdf:ID="attached"/>
    </rdfs:subPropertyOf>
    <rdfs:domain rdf:resource="#MusicSegment"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="atPhysicalTime">
    <rdfs:domain rdf:resource="#MusicSegment"/>
    <rdfs:range rdf:resource="#PhysicalTime"/>
    <rdfs:subPropertyOf rdf:resource="#attached"/>
  </owl:ObjectProperty>
  <owl:ObjectProperty rdf:ID="characterizedBy">
    <rdfs:domain rdf:resource="#MusicSegment"/>
    <rdfs:range>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#TimeLine"/>
          <owl:Class rdf:about="#MusicalConcepts"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:range>
  </owl:ObjectProperty>
  <owl:DatatypeProperty rdf:ID="LAT">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:ID="timbralDescriptor"/>
    </rdfs:subPropertyOf>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="TC">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#timbralDescriptor"/>
    </rdfs:subPropertyOf>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="MPeg7DataTypeDescriptor"/>
  <owl:DatatypeProperty rdf:ID="onTimePoint">
    <rdfs:domain>
      <owl:Class>
        <owl:unionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#MusicSegment"/>
          <owl:Class rdf:about="#TimeLine"/>
        </owl:unionOf>
      </owl:Class>
    </rdfs:domain>
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
    <rdfs:subPropertyOf rdf:resource="#MPeg7DataTypeDescriptor"/>
    <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="HSC">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#timbralDescriptor"/>
    </rdfs:subPropertyOf>
```

```
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="melodicDescriptor">
    <rdfs:subPropertyOf rdf:resource="#MPeg7DataTypeDescriptor"/>
    <rdfs:domain rdf:resource="#MusicSegment"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="hasDuration">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
    <rdfs:domain rdf:resource="#MusicSegment"/>
    <rdfs:subPropertyOf rdf:resource="#MPeg7DataTypeDescriptor"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="HSD">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#timbralDescriptor"/>
    </rdfs:subPropertyOf>
    <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="contourValue">
    <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
    <rdfs:subPropertyOf rdf:resource="#melodicDescriptor"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="HSV">
    <rdfs:subPropertyOf>
      <owl:DatatypeProperty rdf:about="#timbralDescriptor"/>
    </rdfs:subPropertyOf>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:about="#timbralDescriptor">
    <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
    <rdfs:domain rdf:resource="#MusicSegment"/>
    <rdfs:subPropertyOf rdf:resource="#MPeg7DataTypeDescriptor"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="HSS">
    <rdfs:subPropertyOf rdf:resource="#timbralDescriptor"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="AFFT">
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
    <rdf:type
rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty"/>
    <rdfs:subPropertyOf rdf:resource="#melodicDescriptor"/>
  </owl:DatatypeProperty>
  <owl:DatatypeProperty rdf:ID="SC">
    <rdfs:subPropertyOf rdf:resource="#timbralDescriptor"/>
    <rdfs:range
rdf:resource="http://www.w3.org/2001/XMLSchema#float"/>
  </owl:DatatypeProperty>
  <swrl:Imp rdf:ID="MelodyMotionAssociationRule">
    <swrl:body>
      <swrl:AtomList>
        <rdf:first>
          <swrl:ClassAtom>
            <swrl:argument1>
              <swrl:Variable rdf:ID="x"/>
            </swrl:argument1>
            <swrl:classPredicate rdf:resource="#MusicSegment"/>
          </swrl:ClassAtom>
        </rdf:first>
```

```
        <rdf:rest>
          <swrl:AtomList>
            <rdf:first>
              <swrl:ClassAtom>
                <swrl:classPredicate
rdf:resource="#MelodicExpressions"/>
                <swrl:argument1 rdf:resource="#AFFT"/>
              </swrl:ClassAtom>
            </rdf:first>
            <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
          </swrl:AtomList>
        </rdf:rest>
      </swrl:AtomList>
    </swrl:body>
    <swrl:head>
      <swrl:AtomList>
        <rdf:first>
          <swrl:IndividualPropertyAtom>
            <swrl:argument1 rdf:resource="#x"/>
            <swrl:argument2 rdf:resource="#Motion"/>
            <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
          </swrl:IndividualPropertyAtom>
        </rdf:first>
        <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
      </swrl:AtomList>
    </swrl:head>
  </swrl:Imp>
  <Sharpness rdf:ID="AcoustiC"/>
  <Brightness rdf:ID="long"/>
  <Brightness rdf:ID="short"/>
  <Sharpness rdf:ID="CrunchY"/>
  <Sharpness rdf:ID="NasaL"/>
  <swrl:ClassAtom>
    <swrl:classPredicate rdf:resource="#MusicSegment"/>
    <swrl:argument1 rdf:resource="#x"/>
  </swrl:ClassAtom>
  <owl:DataRange>
    <owl:oneOf rdf:parseType="Resource">
      <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
      >-2</rdf:first>
      <rdf:rest rdf:parseType="Resource">
        <rdf:rest rdf:parseType="Resource">
          <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
          >0</rdf:first>
          <rdf:rest rdf:parseType="Resource">
            <rdf:rest rdf:parseType="Resource">
              <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-
rdf-syntax-ns#nil"/>
              <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
              >2</rdf:first>
            </rdf:rest>
            <rdf:first
rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
            >1</rdf:first>
          </rdf:rest>
        </rdf:rest>
        <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#int"
        >-1</rdf:first>
      </rdf:rest>
```

```
      </owl:oneOf>
    </owl:DataRange>
    <swrl:IndividualPropertyAtom>
      <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
      <swrl:argument2 rdf:resource="#Motion"/>
      <swrl:argument1 rdf:resource="#x"/>
    </swrl:IndividualPropertyAtom>
    <Sharpness rdf:ID="RinginG"/>
    <swrl:Imp rdf:ID="SharpnessFeatureAssociationRule">
      <swrl:body>
        <swrl:AtomList>
          <rdf:rest>
            <swrl:AtomList>
              <rdf:first>
                <swrl:ClassAtom>
                  <swrl:classPredicate
rdf:resource="#TimbralExpressions"/>
                  <swrl:argument1 rdf:resource="#HSC"/>
                </swrl:ClassAtom>
              </rdf:first>
              <rdf:rest>
                <swrl:AtomList>
                  <rdf:first>
                    <swrl:ClassAtom>
                      <swrl:argument1 rdf:resource="#HSS"/>
                      <swrl:classPredicate
rdf:resource="#TimbralExpressions"/>
                    </swrl:ClassAtom>
                  </rdf:first>
                  <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-
rdf-syntax-ns#nil"/>
                </swrl:AtomList>
              </rdf:rest>
            </swrl:AtomList>
          </rdf:rest>
          <rdf:first>
            <swrl:ClassAtom>
              <swrl:argument1 rdf:resource="#x"/>
              <swrl:classPredicate rdf:resource="#MusicSegment"/>
            </swrl:ClassAtom>
          </rdf:first>
        </swrl:AtomList>
      </swrl:body>
      <swrl:head>
        <swrl:AtomList>
          <rdf:first>
            <swrl:IndividualPropertyAtom>
              <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
              <swrl:argument1 rdf:resource="#x"/>
              <swrl:argument2 rdf:resource="#Sharpness"/>
            </swrl:IndividualPropertyAtom>
          </rdf:first>
          <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
        </swrl:AtomList>
      </swrl:head>
    </swrl:Imp>
    <swrl:Imp rdf:ID="MelodyShapeAssociationRule">
      <swrl:head>
        <swrl:AtomList>
          <rdf:first>
            <swrl:IndividualPropertyAtom>
              <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
```

```xml
            <swrl:argument2 rdf:resource="#Shape"/>
            <swrl:argument1 rdf:resource="#x"/>
          </swrl:IndividualPropertyAtom>
        </rdf:first>
        <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
      </swrl:AtomList>
    </swrl:head>
    <swrl:body>
      <swrl:AtomList>
        <rdf:rest>
          <swrl:AtomList>
            <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
            <rdf:first>
              <swrl:ClassAtom>
                <swrl:argument1 rdf:resource="#contourValue"/>
                <swrl:classPredicate
rdf:resource="#MelodicExpressions"/>
              </swrl:ClassAtom>
            </rdf:first>
          </swrl:AtomList>
        </rdf:rest>
        <rdf:first>
          <swrl:ClassAtom>
            <swrl:argument1 rdf:resource="#x"/>
            <swrl:classPredicate rdf:resource="#MusicSegment"/>
          </swrl:ClassAtom>
        </rdf:first>
      </swrl:AtomList>
    </swrl:body>
  </swrl:Imp>
  <Sharpness rdf:ID="BrassY"/>
  <Brightness rdf:ID="clean"/>
  <swrl:IndividualPropertyAtom>
    <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
    <swrl:argument2 rdf:resource="#Sharpness"/>
    <swrl:argument1 rdf:resource="#x"/>
  </swrl:IndividualPropertyAtom>
  <Sharpness rdf:ID="Metalic"/>
  <Shape rdf:ID="archedShape"/>
  <Brightness rdf:ID="gentle"/>
  <swrl:ClassAtom>
    <swrl:argument1 rdf:resource="#AFFT"/>
    <swrl:classPredicate rdf:resource="#MelodicExpressions"/>
  </swrl:ClassAtom>
  <Brightness rdf:ID="airy"/>
  <swrl:ClassAtom>
    <swrl:argument1 rdf:resource="#x"/>
    <swrl:classPredicate rdf:resource="#MusicSegment"/>
  </swrl:ClassAtom>
  <Sharpness rdf:ID="HarD"/>
  <swrl:Imp rdf:about="http://www.owl-
ontologies.com/BrightnessFeatureAssociationRule">
    <swrla:isRuleEnabled
rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean"
    >true</swrla:isRuleEnabled>
    <swrl:head>
      <swrl:AtomList>
        <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
        <rdf:first>
          <swrl:IndividualPropertyAtom>
```

```
            <swrl:propertyPredicate rdf:resource="#characterizedBy"/>
            <swrl:argument2 rdf:resource="#Brightness"/>
            <swrl:argument1 rdf:resource="#x"/>
          </swrl:IndividualPropertyAtom>
        </rdf:first>
      </swrl:AtomList>
    </swrl:head>
    <swrl:body>
      <swrl:AtomList>
        <rdf:rest>
          <swrl:AtomList>
            <rdf:first>
              <swrl:ClassAtom>
                <swrl:classPredicate
rdf:resource="#TimbralExpressions"/>
                <swrl:argument1 rdf:resource="#SC"/>
              </swrl:ClassAtom>
            </rdf:first>
            <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-
syntax-ns#nil"/>
          </swrl:AtomList>
        </rdf:rest>
        <rdf:first>
          <swrl:ClassAtom>
            <swrl:argument1 rdf:resource="#x"/>
            <swrl:classPredicate rdf:resource="#MusicSegment"/>
          </swrl:ClassAtom>
        </rdf:first>
      </swrl:AtomList>
    </swrl:body>
  </swrl:Imp>
  <Motion rdf:ID="disjunct"/>
  <Motion rdf:ID="conjunct"/>
  <Shape rdf:ID="rising"/>
  <Sharpness rdf:ID="ResonanT"/>
  <Motion rdf:ID="leaps"/>
  <Brightness rdf:ID="cold"/>
  <swrl:ClassAtom>
    <swrl:argument1 rdf:resource="#HSC"/>
    <swrl:classPredicate rdf:resource="#MusicalConcepts"/>
  </swrl:ClassAtom>
  <Shape rdf:ID="falling"/>
</rdf:RDF>

<!-- Created with Protege (with OWL Plugin 3.4.1, Build 536)
http://protege.stanford.edu -->
```